Identification and Characterization of Rare Variants in Cholinergic Nicotinic Receptor Genes and their Contribution to Substance Dependence

Gabriel Emanuel Haller
Washington University in St. Louis

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

Recommended Citation
Haller, Gabriel Emanuel, "Identification and Characterization of Rare Variants in Cholinergic Nicotinic Receptor Genes and their Contribution to Substance Dependence" (2013). All Theses and Dissertations (ETDs). 1133.
https://openscholarship.wustl.edu/etd/1133

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Identification and Characterization of Rare Variants in Cholinergic Nicotinic Receptor Genes and their Contribution to Substance Dependence

by Gabriel Emanuel Haller

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

August 2013

St. Louis, Missouri
# Table of Contents

Acknowledgements ........................................................................................................ iv

Abstract of the Dissertation .......................................................................................... vi

Chapter 1. Introduction and Perspective ....................................................................... 1

Chapter 2. Identification of rare variants in cholinergic nicotinic receptor genes and their association with nicotine dependence

Summary ........................................................................................................................ 31

Introduction .................................................................................................................. 32

Methods ....................................................................................................................... 34

Results ......................................................................................................................... 38

Discussion ..................................................................................................................... 43

Figures ........................................................................................................................... 46

Chapter 3. Identification of rare variants in cholinergic nicotinic receptor genes and their association with alcohol and cocaine

Summary ........................................................................................................................ 53

Introduction .................................................................................................................. 54

Methods ....................................................................................................................... 56

Results ......................................................................................................................... 59

Discussion ..................................................................................................................... 65

Figures ........................................................................................................................... 69
Chapter 4. Functional characterization of rare variants in \textit{CHRN} \textit{B4} to improve the association between \textit{CHRN} \textit{B4} and nicotine dependence

Summary ......................................................................................................................... 81

Introduction .................................................................................................................... 82

Methods ........................................................................................................................ 85

Results .......................................................................................................................... 93

Discussion ...................................................................................................................... 97

Figures .......................................................................................................................... 101

Chapter 5. Conclusions and Future Directions .............................................................. 159

References ...................................................................................................................... 170
Acknowledgements

First and foremost, I want to thank Dr. Alison Goate, who from the beginning of my time in the lab has treated me more as a peer than an employee. I am so glad I was able to train with her, both because of her wealth of knowledge and countless great ideas, but also for her kindness as a mentor.

I would also like to thank the other members of the Goate lab, past and present, who have made work like home. In particular, my climbing partner, John Budde, without whom no Sequenom would have been done and no crazy idea would have been given hope. I would like to thank the members of my thesis committee, Dr. Rob Mitra, Dr. Joe Henry Steinbach, Dr. John Rice, Dr. Anne Bowcock, Dr. Nancy Saccone and Dr. Michael Bruchas for all of their help over the years and for taking the time to help me grow as a scientist. Much of this work would have been very hard without the help early on of Todd Druley and Francesco Vallania, then both members of the Mitra lab. I am very grateful for their help both devising and performing the pooled sequencing that is at the center of this work. Absolutely none of the functional data presented as part of this dissertation would be possible without the great work of Joe Henry Steinbach, Gustav Akk and Ping Li. I cannot thank them enough.

I would like to of course thank the lead investigators directing data collection: Laura Bierut, Naomi Breslau, Dorothy Hatsukami, and Eric Johnson. In particular, I would like to thank Laura Bierut for all of her help as leader of the COGEND project. The COGEND project is a collaborative research group and part of the National Institute on Drug Abuse (NIDA) Genetics Consortium. Subject collection was supported by NIH grant P01 CA89392 (PI - L Bierut) from the National Cancer Institute (NCI). Phenotypic
and genotypic data are stored in the NIDA Center for Genetic Studies (NCGS) at http://zork.wustl.edu/ under NIDA Contract HHSN271200477451C (PIs J Tischfield and J Rice). Electrophysiological studies supported by NIH grant R21 DA26918 (G. Akk) and NINSD grant R01 NS22356 (J.H. Steinbach).

The COGA project is a collaborative research group and part of the National Institute on Drug Abuse (NIDA) Genetics Consortium. Subject collection was supported by NIH grant XX (PI - L Bierut) from the NIXX. Phenotypic and genotypic data are stored in the NIDA Center for Genetic Studies (NCGS) at http://zork.wustl.edu/ under NIDA Contract HHSN271200477451C (PIs J Tischfield and J Rice).

I would very much like to thank my family. My parents and my sisters, as well as my entire extended family, have always supported my desire to be a scientist and a geneticist. They have never tried to tell me what to be, just that I should do what I love.

Lastly, I thank my fiancée Brooke Sadler. Thank you for everything!
ABSTRACT OF THE DISSERTATION

Identification and Characterization of Rare Variants in Cholinergic Nicotinic Receptor Genes and their Contribution to Substance Dependence.

by

Gabriel Emanuel Haller

Doctor of Philosophy in Biology and Biomedical Sciences
Human and Statistical Genetics Program
Washington University in St. Louis, 2013

Professor Alison Goate, Chairperson

Genome-wide association studies (GWAS) have identified common variation in the CHRNA5-CHRNA3-CHRNB4 and CHRNA6-CHRNB3 gene clusters that contribute to nicotine dependence. However, the role of rare variation in risk for substance dependence in these nicotinic receptor genes has not been studied.

In order to investigate the role rare variants play in the development of nicotine dependence, we undertook pooled sequencing of the coding regions and flanking sequence of the CHRNA5, CHRNA3, CHRNB4, CHRNA6 and CHRNB3 genes in 710 African Americans (461 nicotine dependent (ND) cases and 249 smokers with no symptoms of dependence (controls)) and 2055 European Americans (1062 ND cases and 993 controls) from the Collaborative Study of the Genetics of Nicotine Dependence (COGEND). Carrier status of individuals harboring rare non-synonymous variants at conserved sites in each of these genes was then compared in cases and controls to test for an association with nicotine dependence. We find a reduced risk for nicotine dependence among carriers of missense variants at conserved residues in CHRNB4 in African Americans and European Americans.
We next aimed to determine whether rare genetic variation in these genes influence risk for developing alcohol or cocaine dependence, two conditions highly comorbid with nicotine dependence. We undertook pooled sequencing of the coding regions and flanking sequence of the \textit{CHRNA5}, \textit{CHRNA3}, \textit{CHRNB4}, \textit{CHRNA6} and \textit{CHRNB3} genes in 287 African Americans (147 DSM-IV alcohol dependent cases and 140 controls) and 1028 European Americans (480 DSM-IV alcohol dependent cases and 548 controls) individuals from the Collaborative Study of the Genetics of Alcoholism (COGA). For European Americans, we find increased DSM-IV cocaine dependence symptoms (famSKAT p= $2 \times 10^{-4}$) and increased DSM-IV alcohol dependence symptoms (famSKAT p= $5 \times 10^{-4}$) among carriers of missense variants in \textit{CHRNB3}. For African Americans, we find decreased cocaine dependence symptoms among carriers of missense variants in \textit{CHRNA3} (famSKAT p= 0.006).

Finally, we sought to determine the functional impact of rare variants in \textit{CHRNB4} harbored by individuals in these two cohorts. A total of 10 variants in \textit{CHRNB4} observed in the COGEND cohort were investigated. One variant in \textit{CHRNA3} was also functionally tested as it is high linkage disequilibrium with another rare missense variant in \textit{CHRNB4}. We find several variants that alter cellular response either to nicotine or acetylcholine as well as many variants that alter cell-surface protein expression as measured by cell-surface ELISA without altering either \textit{CHRNB4} mRNA or total β4 protein as measured by western blot. Further, when we integrate these \textit{in vitro} findings into a model of association with nicotine dependence related traits, we improve the association, suggesting that the success of future association analyses in these and potentially many...
other genes across the genome, may depend greatly on functional assessment of observed
genetic variation.
Chapter 1. Introduction and Perspective

*Complex Trait Genetics*

A complex genetic trait is defined as a heritable trait for which the genetic contributors are not limited to one or a small number of genes. This is in stark contrast to what geneticists refer to as “Mendelian” traits or diseases for which the genetic causes are comparatively simple, being limited to one or a handful of genes or loci. Much effort in the preceding decades has been spent trying to understand the genetic underpinnings of complex traits. Before the advent of inexpensive single nucleotide polymorphism (SNP) genotyping microarrays, limited numbers of variable sites across the genome were interrogated either to ask very specific questions about genes with high prior probabilities of contributing to a trait or to perform linkage scans in families to identify large genomic regions segregating with disease/trait status. In each case, genes underlying complex traits were identified, but with limited success. More recently, microarrays able to genotype hundreds of thousands or millions of SNPs across the genome have been used to identify thousands of common genetic variants contributing to complex disease risk. So-called genome-wide association studies (GWAS) were then performed using these genotyping arrays in large cohorts of cases, individuals with a given disease or trait, and controls, individuals specifically ascertained to be free of disease or not have a given trait. The variants significantly associated with the tested diseases or traits have generally had small effects, however, with individuals heterozygous at these sites having

---

odds of developing disease of 1.1-1.5 compared to non-carriers. These small effect-sizes have demanded that investigators increase sample size considerably in order to have the statistical power necessary to find significant associations. One of the largest studies of a complex trait, consisting of ~250,000 subjects, was for anthropomorphic traits such as height and body-mass index (BMI; height/weight\(^2\)) \(^2\). This study was able to find 69 genome-wide significant SNPs associated with one of their traits of interest but were still only able to explain ~5\% of BMI with the genome-wide significantly associated SNPs (Figure 1.1).

More recent studies of complex traits have utilized the newly developed method of exome sequencing, i.e. only sequencing the protein-coding portion of the genome (~1.5\% of the genome)\(^3\)\(^{-}\)\(^5\) or have used genotyping microarrays designed to only interrogate the protein-altering genetic variants (exome chip). Exome sequencing involves using genomic capture techniques, i.e. hybridizing sheared genomic DNA with biotinylated RNA baits complementary to the protein-coding regions of the genome and then washing away any DNA fragments that did not hybridize (Figure 1.2). For the exome-chip, only protein-altering variants are genotyped that were found in multiple individuals from ~12,000 sequenced exomes and genomes culled from the 1000 genomes project\(^6\), the NHLBI Exome Sequencing Project\(^7\) and others. In each case, the population frequency of variants is generally much lower than the frequency of variants investigated as part of genome-wide association studies, as protein-altering variants are more likely to be negatively selected in populations due to their increased probability of deleterious effects and as a result are also more likely to be at low frequencies in populations\(^8\). This reduced population frequency, however, also reduces power to detect associations between
complex traits and these variants. As such, very few of the studies utilizing these methods have had much success identifying novel genes or loci underlying complex disease. Even using collapsing methods wherein rare variants are combined to create pseudo-common variants or using complex statistical frameworks, very few examples exist of large exome-sequencing efforts that have yielded significant findings and many of these have been for de novo variants in schizophrenia, autism and cancer. As exome sequencing is still rather expensive, however, many of the exome sequencing studies have been made of many small samples pieced together from disparate sources and as such have lacked power. Regarding studies utilizing the exome-chip, the lack of positive results may be in part due to the fact that rarer variants are harder to call using currently available genotype clustering algorithms. The result of this is an increased probability of incorrect genotyping calling which in turn increases both the false positive and false negative rates in association testing done with the resultant genotypes. Very few studies, therefore, have yet been published using exome-chip data.

Overall, complex genetic diseases have garnered substantial attention in the recent past, but only small amounts of variability in the traits and diseases under study have been explained using current techniques. This can be explained in part by a lack of attention paid to rarer variants and in part due to a lack of power in all studies. It is likely that complex traits and diseases are the product of hundreds if not thousands of variants of all frequencies, many with near infinitesimal effect sizes. It will likely be important in the future to develop ways of integrating knowledge of specific variant’s impact on protein function or gene expression, etc. into models of association.
Substance abuse is one the leading causes of preventable death in the United States and across the world. Cigarette smoking, for instance, dramatically increases risk of developing lung cancer, chronic obstructive pulmonary disease (COPD), coronary heart disease and stroke and is overall the number one cause of preventable death worldwide. Figure 1.3 shows survival curves for individuals who either never smoked, used to smoke or smoke differing numbers of cigarettes per day on average. Additionally, alcohol dependence represents a considerable health and economic burden due to substance use related morbidity, accidents and imprisonment. As such, it is critical to understand the underlying causes of substance-related behavior and substance dependence (SD). Studies have suggested that genes, the environment and the interplay between the two play an important role in determining whether or not a person will become substance dependent. For instance, an individual who has never been exposed to nicotine cannot be nicotine dependent. If one is concerned with elucidating genes underlying the effect of nicotine on smoking behavior, it is important to correct for environmental factors whenever possible. It is clear from epidemiological studies and in ascertaining individuals for genetic studies that there exist different strata of smokers. There are those who have smoked cigarettes, but not enough to elicit sustained activation of reward pathways in the brain (~22% of smokers), those who smoke regularly but show few signs of being nicotine dependent (~33% of smokers) and finally those who show high levels of nicotine addiction (~44% of smokers). Figure 1.4 shows the proportions of individuals from a population-based sampling of individuals who made the various
transitions to nicotine dependence and the heritability estimates for both smoking initiation and nicotine dependence\textsuperscript{22}. Additionally, because there is strong comorbidity among types of substance dependence and it is thought that though there are substance specific genetic and environmental risk factors, some evidence suggests that much of the genetic vulnerability to different substances is shared\textsuperscript{23,24}.

Using the Fagerström Test for Nicotine Dependence (FTND), a quantitative measure of nicotine dependence can be calculated\textsuperscript{25,26}. The FTND questionnaire consists of six questions (4 one point questions and 2 three point questions) giving a maximum score of 10 for individuals with strong nicotine dependence (Figure 1.5). A major component of the FTND is the number of cigarettes smoked per day (CPD), one of the two three point questions, making CPD another quantitative measure of nicotine dependence similar to and highly correlated with FTND score. Additionally, we can, I we do for our studies, exclude all individuals who have never smoked at least 100 cigarettes in their life. This corrects for the GxE interaction between nicotine exposure and associated gene variants with respect to nicotine dependence. It is for these reasons that we will choose individuals to sequence from the extremes of FTND score and CPD.

Studies of families and twins have suggested a strong genetic contribution to the development of multiple nicotine related traits. One such study found that \textasciitilde 50\% of the variance in smoking persistence could be accounted for by genetic factors\textsuperscript{27}. Additionally, age of initiation was also found to be highly heritable\textsuperscript{28}. A meta-analysis of several twin studies also identified a role for genetics in the initiation of smoking\textsuperscript{29}. A number of heritability studies have investigated generalized drug addiction, often including nicotine dependence as one of the drugs on which an individual could be
dependent. One such study found that there was as much as an 8-fold increase in risk for relatives of drug-dependent individuals compared to individuals without drug-dependent relatives. Twin studies have also given credence to the genetic contributions to smoking behavior. Many groups have found that the concordance for smoking was significantly higher in monozygotic male twin pairs than in dizygotic male twins.

Genetic variation in the cholinergic nicotinic receptor genes (CHRNs) has repeatedly been found to be highly associated with nicotine dependence (see for review). The first report of variants in nicotinic receptor genes contributing to nicotine dependence was a candidate gene study performed by members of our group. The most compelling association from that study was with a SNP near CHRN3 on chromosome 8 in humans. However, they were also the first to report an association between the non-synonymous SNP rs16969968 (D398N in the a5 nicotinic receptor subunit gene). Despite big differences in the frequency of this SNP across populations, it shows a similar effect on risk, with odds ratios 1.3-2.00 in European, African American and Asian populations.

Notably, nicotinic receptor genes often form gene clusters. CHRNA5, CHRNA3 and CHRN4 on chromosome 15 and CHRNA6 and CHRN3 on chromosome 8 are directly adjacent to each other (Figure 1.6). This is of particular note, as at least some of these genes seem not to have evolved from tandem duplications, but rather most likely came together having been duplicated elsewhere to exist in close proximity on one haplotype block so that they could be co-regulated. If a phylogenetic tree is constructed using the amino acid sequences of all human nicotinic receptor paralogs, it is clear that the alpha subunits are more closely related to each other than to any beta subunit and that CHRNA3
shares more similarity to other alpha subunits than to \textit{CHRNA5}, its closest physical neighbor (Figure 1.7). The hypothesis that these genes are co-regulated is confirmed in part when one considers the construction of “tabak” mice, constructed to possess a transgenic region encompassing all of the genes in the cluster, including non-coding regions \footnote{Reference}. These mice overexpress \textit{chrnb4} and express an eGFP-\textit{chrna3} hybrid protein in precisely the regions of the brain that express the genes endogenously, suggesting that the surrounding non-coding DNA alone is sufficient to determine brain expression patterning. Since the discovery of the role of rs16969968 in nicotine dependence, multiple independent variants near the \textit{CHRNA5} gene have been identified that contribute to smoking related phenotypes. The discovery of these variants was in part motivated by the presence of other nicotinic receptor genes in the region of association. A group of highly correlated variants including the SNP rs588765 near \textit{CHRNA5} was recently shown to increase \textit{CHRNA5} mRNA expression and increase risk of nicotine dependence independently of rs16969968 \footnote{Reference}. Additionally, a group of SNPs tagged by rs1996371 near the \textit{CHRNB4} gene was shown to affect age of onset of daily smoking \footnote{Reference}. Of note, this variant is not in high linkage disequilibrium ($r^2$) with either the common non-synonymous variant in \textit{CHRNA5} (rs16969968) or the variants correlated with \textit{CHRNA5} mRNA (tagged by rs588765), suggesting as many as three independent associations between smoking related phenotypes and variants in this region. The common variants at the \textit{CHRNA6-CHRNB3} gene cluster as well as the common variants at the \textit{CHRNA5-CHRNA3-CHRNB4} gene cluster tagged by rs16969968 were also recently shown to contribute to smoking quantity in several large meta-analysis genome-wide association studies \footnote{Reference, Reference}. Together these results strongly suggest multiple mechanisms connecting
nicotinic receptor function to nicotine dependence and nicotine related behaviors. Despite these findings, only a small proportion of the variance (~5%) in nicotine dependence related traits has been explained by these variants 45.

The same genome-wide association study meta-analysis that identified variants in the CHRNA6-CHRN3 and CHRNA5-CHRNB3 gene clusters, also identified a group of variants near the CYP2A6 gene (tagged by rs4105144) that were genome-wide significantly associated with cigarette consumption 43. The CYP2A6 gene was an ideal candidate for such a study as it is the major hepatic nicotine-metabolizing enzyme in humans. There are several variants in or near CYP2A6 that reduce enzymatic activity and others that are highly associated with rate of nicotine metabolism 44,46. The GWA associated SNP is in LD with CYP2A6*2 (rs1801272) (r2 = 0.13 and D’ = 1.0 in the HapMap CEU samples) and the CYP2A6*2 reduced-function allele is only found on the background of the minor allele of rs4105144, associated with reduced smoking quantity. Further, our group has demonstrated that an allele defined by harboring a large deleted segment containing this SNP is also in complete LD (D’=1) with the GWA associated SNP. This not only suggests that the SNP rs4105144 is tagging several reduced-function variants, but that its association may have been partially exaggerated by having some hemizygous individuals who are heterozygous for rs4105144 appear as minor allele homozygotes.

*Rare Variant Associations*

Rare genetic variants have recently been shown to contribute to a number of common human diseases (see 47 for review). Multiple rare variants have been reported in
genes previously shown to harbor common variants associated with common diseases\textsuperscript{48-50}. Additionally, studies have demonstrated associations between rare variation in nicotinic receptor genes and nicotine dependence\textsuperscript{51,52}. As a class, rare SNPs, copy-number variants (CNVs) and small insertion/deletion polymorphisms (indels) constitute the majority of human genetic variation and thus may hold the key to understanding part of the missing heritability of complex traits unaccounted for by recent genome-wide association studies. Sequencing is one of the few methods by which these variants can be investigated; however, the cost of sequencing hundreds or thousands of individuals for multiple genes remains high despite advances in sequencing technologies. An efficient alternative approach is to carry out DNA sequencing using DNA pooled from multiple individuals. Alternatively, whole exomes have recently become a relatively efficient method of performing gene-based, rare variant associations in an unbiased manner across the genome. This technique was first utilized for the identification of rare Mendelian diseases, but has recently expanded into gene-based genome-wide searches for common complex diseases. One of the first Mendelian disorders to be mapped using this method was Kabuki syndrome\textsuperscript{4}. The identification of the gene underlying this disorder was relatively straightforward, however, as it is recessive and thus required each affected individual to be either homozygous or a compound heterozygote for the same gene, a condition that is not met with high probability in the general population. Efforts have been far less successful for common complex, however, likely due to the limited power afforded by small sample size when using collapsed non-synonymous variants. Such variants are exceedingly rare, especially in genes with even moderate levels of cross-species conservation, and as such require large sample sizes to obtain sufficient power to
detect an association. Candidate gene studies have had more success than whole-genome scans, likely due to the reduced burden of multiple testing. A study of five genes requires a modest p-value of 0.01 to reach statistical significance, while an exome sequencing study requires a p-value of $2.5 \times 10^{-6}$ assuming 20,000 genes to reach statistical significance.

Methods for the analysis of multiple rare variants with regards to complex traits are still in their infancy. A number of studies have suggested strategies for investigating rare variants. Among these are the cohort allelic sums test (CAST), combined multivariate and collapsing (CMC), weighted sum (WS), c-alpha, variable threshold (VT) and SKAT methods. Each technique combines multiple alleles to create aggregate genotypes, scores or linear models to be used in a test of association. In the case of the CAST method, the number of individuals with at least one mutation at a locus is compared in cases and controls. This is by far the simplest method, but has limitations. For the CMC method, rare variants are collapsed and then used with all common ($\geq 5\%$) variants in a multivariate analysis. One of the more intriguing techniques developed thus far for detecting rare variant associations is the weighted sum method. In this method, variants are weighted in proportion to the inverse of their frequency and each weighted variant present in an individual is summed to form a person’s score. The scores of cases and controls are then compared by permutation. One aspect lacking in each of the aforementioned methods is some measure of the likelihood of functional impact for each of the variants used in the analysis. For instance, evolutionary conservation, regional genetic diversity or an actual measure of function for each variant would likely improve power if integrated into one of these analyses. One of the largest problems facing human
geneticists is how to determine what genetic variation present in a population is functional, in particular whether variants are functional with respect to the phenotype under study. The weighted sum method implicitly claims a relationship between variant frequency and the probability that the variant has functional consequences. This is a valid claim as far as sites under negative selection will tend to have reduced frequency compared to neutral sites. However, some variants are rare because they entered the population only a short time ago; they have not had sufficient time to be acted upon by either genetic drift or natural selection. A more direct estimate of a site's probability of function is its level of evolutionary conservation. Conservation is a proxy for functional constraint and the presence of negative selection. I have previously described a method using Tajima’s D, a measure of genetic diversity compared to divergence, to compare case and control allele frequency spectra at a locus. Another possibility would be to weight variants as in the weighted sum method but by likelihood ratio test (LRT) score, a measure of conservation, rather than by frequency. Recently, progress in the field of rare variant association testing was made with the introduction of the sequence kernel association test (SKAT) and the c-alpha test. In the case of SKAT, the test is a score-based variance-component test that takes into account both rare and common variants. One limitation of collapsing methods is that if neutral common alleles are collapsed with even rare alleles with large effects, the common variant will mask the effects of the rare alleles. This is overcome at least in part by SKAT in that alleles are tested independently within the framework. For the c-alpha test, the distribution of allele frequencies is compared between cases and controls. Whether rare alleles occur more frequently in cases or controls, the distribution of alleles between cases and controls is
skewed and will be identified. Despite the success these methods have afforded researchers in the field, further work is necessary to perfect tests of rare variant association and improve our ability to detect these rare variants often with modest effect sizes.

Recently, several groups have interrogated the contributions of rare genetic variants in select nicotinic receptor genes. Specifically, one group sequenced the CHRNA2, CHRNA3, CHRNA4, CHRNA5, CHRNA6, CHRNA7, CHRN1, CHRN2, CHRN3, and CHRN4 genes in ~1200 European Americans and another sequenced the CHRNA4 gene in ~200 ND cases and ~200 controls and the fifth exon of CHRNA4 in around 1000 additional ND cases and 1000 additional controls. Together, these groups were only able to observe modest associations between nicotine dependence and variants, either singly or collapsed across rare variants. They may have been limited by having used only European Americans, however, as African Americans are known to harbor a larger proportion of rare variation as they did not go through a recent population bottleneck. These results highlight the need for more thorough investigation of rare genetic variation in these nicotinic receptor genes with respect to susceptibility to nicotine dependence.

Genetics of Alcohol and Cocaine Dependence

Twin and adoption studies have suggested that in addition to genetic variants that contribute to risk of becoming dependent on a specific drug, there exist genetic factors underlying a generalized risk for becoming dependent on multiple substances (see for review). Two such studies compared the rates of drug use for several commonly used
substances over time in a cohort of monozygotic (MZ) and dizygotic (DZ) twins, finding that the strong correlation observed between several drug dependence phenotypes could be accounted for in part by shared genetic factors\textsuperscript{59,60}. The findings suggest that there likely exist variants that contribute to risk for multiple drug dependence phenotypes independently and that it is therefore possible that variation in nicotinic receptor genes, despite having a clear mechanism of action with respect to nicotine dependence, may contribute to risk for other drug dependences.

Epidemiological studies have shown high levels of comorbidity among alcohol, cocaine and nicotine dependence\textsuperscript{23}. As such, it has been difficult to disentangle the effects of genetic variants on each of these phenotypes. Numerous studies have found that variants altering the function or expression of neuronal cholinergic nicotinic receptors (\textit{CHRN}s) alter risk for becoming nicotine dependent and several have shown effects of variants in nicotinic receptors on risk for both alcohol and cocaine dependence (see\textsuperscript{35} for review). It is unclear from some of these studies, however, how much of the association with a specific phenotype is due to the effect of the variant on nicotine versus alcohol or cocaine. Common single nucleotide polymorphisms (SNPs) in \textit{CHRNA5} and \textit{CHRNA6} have been shown to be associated with nicotine dependence or nicotine consumption and the most strongly associated SNP in several GWAS of nicotine dependence, a non-synonymous change (rs16969968/D398N) in \textit{CHRNA5}, has been shown to also affect risk for cocaine dependence\textsuperscript{61,62}. Additionally, a highly correlated group of variants near \textit{CHRNA5} previously shown to alter \textit{CHRNA5} mRNA expression \textit{in vivo}, alter risk for alcohol dependence\textsuperscript{41}. A group of common SNPs near the \textit{CHRNA3-CHRNA6} gene cluster were also shown to affect cigarette consumption in a recent GWAS\textsuperscript{43}. Together,
common variants explain only a small proportion of the variance (~1%) in alcohol or cocaine dependence \(^4^5\). The Diagnostic and Statistical Manual of Mental Disorders Version IV (DSM-IV) is commonly used to assess and define alcohol dependence. The DSM-IV defines alcohol dependent cases as individuals who endorse three (or more) of the 7 symptoms of alcohol dependence at any time in the same 12-month period. Using these criteria, we and others have identified a number of loci across the genome either implicated in susceptibility to alcohol by linkage or by association. Among these are several variants in gamma-aminobutyric acid receptors (\(GABRA\)), alcohol dehydrogenase (\(ADH\)), aldehyde dehydrogenase (\(ALDH\)), the mu opioid receptor (OPRM1) and bitter taste receptors (TAS) genes (see \(6^3\) for review). Only two loci have been implicated in alcohol dependence at the genome-wide significant level (< \(5 \times 10^{-8}\)): A missense variant (rs1229984; R48H) in the \(ADH1B\) gene and a bin of common SNPs tagged by rs12912251 near C15orf53 \(6^4\). Most of the other loci have only been weakly implicated as risk factors for alcohol dependence.

As for alcohol dependence, several candidate loci have been implicated in risk for cocaine dependence either by linkage or association studies \(^6^5-^7^3\). Among these are two opioid receptor genes (\(OPRM1\) and \(OPRD1\)), genes involved in serotonin and dopamine systems like the 5-hydroxytryptamine (serotonin) receptor 1B gene (HTR1B) or the dopamine receptor D2 (DRD2), as well as a number of other candidate genes. Unlike alcohol dependence, no variant to date has been genome-wide significantly associated with cocaine dependence \(^7^4\). This is because no genome-wide association study of cocaine dependence has been performed. Additionally, it is unclear how many of these genetic factors are contributing to cocaine dependence specifically or whether these
variants are affecting risk through their effects on susceptibility to other comorbid traits or on a generalized effect on drug dependence.

**Functional Studies of Nicotinic Receptors**

Neuronal nicotinic acetylcholine receptors are pentameric ion channels produced from various combinations of α and β receptor subunits and are the main target for nicotine in the brain (Figure 1.8-1.10). The vast majority of neuronally expressed nicotinic receptors contain the α4 and β2 subunits, but many other combinations are formed and have distinct properties and expression patterns. Figure 1.8 shows the distribution of various nicotinic receptor subunit combinations in the mouse brain. Many studies have interrogated the effects of altering either nicotinic receptor amino-acid sequence or expression level. Several nicotinic receptor genes have been characterized in mice by knocking out the gene or transgenic overexpression and several variants observed in human populations have been cloned and overexpressed to determine their effects using electrophysiological methods or calcium flux assays. To date, the Chrnb2, Chrnb4, Chnra6, Chrna4, Chrna7, Chrna3 and Chnra5 genes have been knocked out and the *CHRNB4* gene has been overexpressed in mice. There are profound effects on nicotine-induced behaviors in each case, strengthening the now well-established role of nicotinic receptors in determining the physiological effects of nicotine. The specifics regarding the role of each nicotinic receptor subunit are quite wide-ranging, however, though there are commonalities. For instance, both the deletion of Chrna5 and the deletion of Chrnb4 results in resistance to nicotine induced seizures. At extremely high doses of nicotine, mice will enter an epileptic state. Mice lacking either of these genes can be administered
doses far exceeding the dose required to induce seizures in wild-type mice without noticeable effects. Many other drug related and behavioral changes are apparent in mice either lacking or overexpressing neuronal nicotinic receptor subunits. While knockout mice for $\beta_2$, $\alpha_4$ or $\alpha_6$ nicotinic subunits did not self-administer nicotine, there is no difference between $\alpha_7$ knockout mice and wild-type as far as nicotine self-administration.

In contrast, mice lacking Chrna5 demonstrate enhanced nicotine self-administration. With regard to drug independent behavioral changes, several nicotinic receptor knockout lines show differences from wild-type mice. For instance, locomotor behavior can be assessed in an open field test and can be categorized as being more navigational or more exploratory. Using this test, $\beta_2$ knockouts show a shift toward navigation to the detriment of exploration. Re-expressing this subunit in the ventral tegmental area via lentiviral rescue was able to normalize this phenotype, suggesting the role of $\beta_2$ subunits in this brain region in balancing these two aspects of locomotor behavior.

A number of in vitro studies of variants in nicotinic receptor genes have been performed in order to assess the functional impact of genetic variants on the receptors harboring them. The three major techniques for performing these assays are: 1) voltage-clamp of HEK cells transiently or stably transfected with expression vectors containing wild-type or mutant nicotinic subunits, 2) voltage-clamp of xenopus oocytes injected with either wild-type or mutant mRNA and 3) calcium flux assays that utilize aequorin, a composite of the apoprotein apoaequorin and the prosthetic group coelenterazine, to assess agonist induced calcium influx. One such assay was used to demonstrate that receptors including the amino acid encoded by the minor/risk allele of the SNP rs16969968 ($CHRNA5$ D398N) show a reduced response to the nicotinic agonist,
epibatidine \textsuperscript{85}. These findings are in contrast, however, to results from whole-cell patch clamp of HEK293 cells stably expressing α3 and β4 nicotinic receptor subunits, transiently transfected with either the 398D or 398N form of the α5 nicotinic subunit \textsuperscript{86}. This study was not able to observe a significant difference between the normal and variant form of the receptor. Recent findings from our group have suggested that a high extra-cellular concentration of calcium is necessary to observe the effect, a condition met as a requirement of the experimental setup for the calcium flux assay used to demonstrate the functional effect of the SNP initially \textsuperscript{87}.

Several variants in the \textit{CHRNB4} gene have been shown to affect receptor function using one of these \textit{in vitro} assays \textsuperscript{88}. The \textit{CHRNB4} variants T91I, R136W, S140G, and M467V were each expressed heterologously in Xenopus oocytes and studied using the two-electrode voltage clamp. These experiments revealed that the 136W and 467V forms of the receptor had higher sensitivity to acetylcholine and lower EC\textsubscript{50} than the wild-type. The T91I variant was shown to have lower sensitivity to acetylcholine and a larger EC\textsubscript{50}. Each of the T91I, R136W, and M467V variants also showed greater desensitization after long-term exposure to low-dose acetylcholine.

One study of human carriers of the \textit{CHRNA5} D398N variant using resting-state functional MRI found that carriers of the 398N allele had decreased intrinsic resting connectivity strength in the dorsal anterior cingulate–ventral striatum/extended amygdala circuit \textsuperscript{89}. Though, as mentioned before, the vast majority of neuronally expressed nicotinic receptors contain the α4 and β2 subunits, many other combinations are formed and have distinct properties and expression patterns. It is thought that the various aspects of nicotine dependence (craving, withdrawal, tolerance, sensitivity, etc.) may be due to
these distinct characteristics. For example, recent studies have shown that mice lacking Chrna5 fail to display the adverse effects of high doses of nicotine and mice lacking Chrnb4 or Chrna5 show reduced somatic signs of withdrawal after ceasing long-term nicotine treatment. Additionally, mice over-expressing Chrnb4 were recently shown to display a strong aversion to nicotine, that can be reversed by virally mediated expression of α5 D398N variant in the medial habenula. This work suggests that the aversive effects of nicotine in mouse are regulated by the balanced activity of β4 and α5 nicotinic receptor subunits in the medial habenula. A recent association study of smoking cessation also suggests that variants in CHRNA4 may decrease craving and withdrawal symptoms. These findings suggest that multiple nicotinic receptor subunit genes likely play a role in the development and maintenance of nicotine dependence and that variants in these genes may have different and in some cases opposing functional effects in humans and mice.

In addition to mediating the effects of nicotine, neuronal nicotinic receptors have also been shown to alter the cellular and behavioral effects of alcohol independently of nicotine. In vitro studies have shown altered response to ethanol upon perturbation of nicotinic receptors in neurons. Studies in rats have extensively shown that partial agonists of α4β2* or α3β4* nicotinic receptors affect alcohol consumption in vivo. These results were confirmed and expanded using transgenic and knockout mice of various nicotinic receptor genes. Further, one recent human trial found that the α4β2 subunit partial agonist Varenicline was able to reduce alcohol consumption among smokers. Together these results suggest that altered nicotinic receptor function...
likely has an effect on alcohol consumption independent of its effect on nicotine consumption.

Conclusions

Overall, work on the genetic basis of substance dependence, in this case nicotine, alcohol and cocaine dependence, has provided evidence that variants in the neuronally expressed nicotinic receptor genes can contribute to risk for these substance dependences. Work from many in the field has demonstrated that rare variation in genes may be able to explain a substantial proportion of the heritability still missing despite massive meta-analysis efforts to uncover common variants underlying common disease. Together, these results suggest that interrogation of the role of rare coding variation in the neuronal nicotinic receptor genes will likely aid our understanding of the role of these genes in the etiology of substance dependence. In the following chapters, I will describe how we undertook sequencing of the most promising nicotinic receptor genes, those previously implicated by genome-wide association studies to be involved in nicotine dependence, to characterize the role of rare protein-altering variation in substance dependence risk. We hypothesize that rare variants in the 5 nicotinic receptor genes contained within the two gene clusters previously implicated in genome-wide associations studies, \textit{CHRNA5, CHRNA4, CHRNA3, CHRNA6 and CHRNA3}, harbor rare coding variation associated with substance dependence, specifically nicotine, alcohol and cocaine dependence. Furthermore, as work to improve our ability to detect rare variant association has suggested that our ability to differentiate between protective, neutral and risk alleles will substantially improve our ability to both find and understand rare variant associations, I will describe how we functionally characterized all observed protein-altering variation in
**CHRNB4** to investigate incorporating the observed functional impact of specific variants could improve our understanding of how these rare variants are affecting substance dependence risk. We hypothesize that incorporation of such information will enhance our association signal.
Figure 1.1. Variance Explained at Differing uncorrected P-value Thresholds. Figure is taken from 2. A) Studies using only the extremes of BMI B) Studies using BMI as a quantitative trait. Solid line is variance explained at each level of significance in the meta-analysis of all listed constituent studies.
Figure 1.2. Genomic Capture Strategy. Genomic DNA is prepared and hybridized with a library composed of biotinylated RNA complementary to regions of the genome to be captured. DNA-RNA hybrids are then incubated with streptavidin coated magnetic beads. All DNA not hybridized with the biotinylated-RNA is then washed away. The retained DNA is then amplified and sequenced. Figure is taken from: http://www.genomics.agilent.com/.
Figure 1.3. Survival of men who either never smoked, are ex-smokers or smoked varying numbers of cigarettes per day on average. For example, ~65% of men who smoked ≥ 25 cigarettes per day lived to be 65 years old. Figure is taken from 18.
Figure 1.4. Proportion of individuals(160,214),(826,797) transitioning from various steps in the progression to nicotine dependence. The heritabilities of the smoking initiation (left) and nicotine dependence (right) are shown in red. Dependence is based on the FTND scale. Data is from\textsuperscript{20,85}. 
Figure 1.5. Fagerström Test of Nicotine Dependence. Each question is worth 1 or 3 points and the sum of the points assigned based on answers provided are summed to form a score from 1-10. A score above 4 confers a diagnosis of nicotine dependence. Information is from 25,26.
Figure 1.6. Genomic context of sequenced genes. Positions are from human genome reference assembly build 36.1 (hg18). Direction of transcription is designated with an arrow. Exons are show as squares.
Figure 1.7. Phylogenetic tree of nicotinic receptor cDNA sequences. Two subunits are in a cluster on chromosome 8 (red) and three subunits form a cluster on chromosome 15 (green). The most abundant subunit combination is highlighted in orange and muscle alpha and beta nicotinic receptor subunits are highlighted in blue. Figure was created using the ClustalW program \footnote{117}.
Figure 1.8. Regional distribution of nicotinic acetylcholine receptors. (a) Subunit combinations were deduced either by binding, immunoprecipitation and/or immunopurification assays or from in situ hybridization, single-cell PCR and binding studies of tissues obtained from rat and/or wild-type and/or knockout mice. * refers to any of many possible subunits. Figure is taken from 118.
Figure 1.9. Cytisine resistant $^{125}$I-epibatidine binding. Slices of brains from mice overexpressing chrnb4 were incubated with cytosome and subsequently subjected to $^{125}$I-epibatidine. Darker areas correspond to increased epibatidine binding not bound by unlabeled cytisine, which preferentially binds alpha4-beta4 containing receptors. Figure is taken from 39.
Figure 1.10. Nicotinic acetylcholine receptor configuration within the cell membrane (green). Subunits (red) form pentameric ion channels. Upon binding of acetylcholine (ACh) or other agonists, the pore is opened and the cell is depolarized. Binding of agonist (yellow circles) occurs at the interface between alpha and beta subunits in the case of heteromeric receptors. Figure is taken from 75.
Chapter 2. Identification of rare variants in cholinergic nicotinic receptor genes and their association with nicotine dependence

ABSTRACT

Genome-wide association studies (GWAS) have identified common variation in the \textit{CHRNA5-CHRNA3-CHRNB4} and \textit{CHRNA6-CHRNB3} gene clusters that contribute to nicotine dependence. However, the role of rare variation in risk for nicotine dependence in these nicotinic receptor genes has not been studied. We undertook pooled sequencing of the coding regions and flanking sequence of the \textit{CHRNA5, CHRNA3, CHRNB4, CHRNA6} and \textit{CHRNB3} genes in African American and European American nicotine dependent smokers and smokers without symptoms of dependence. Carrier status of individuals harboring rare missense variants at conserved sites in each of these genes was then compared in cases and controls to test for an association with nicotine dependence.

We find a reduced risk for nicotine dependence among carriers of missense variants at conserved residues in \textit{CHRNB4} in African Americans and European Americans (AA \textit{p}=0.0025, OR=0.31, 95\% CI=0.31-0.72; EA \textit{p}=0.023, OR=0.69, 95\% CI=0.50-0.95). We note that given the possibility of stochastic differences in rare allele frequencies between groups replication is necessary to confirm these findings.

---

INTRODUCTION

Genetic variation in the cholinergic nicotinic receptor genes (CHRN)s has repeatedly been found to be highly associated with nicotine dependence (see 35 for review). Multiple independent variants have been identified that contribute to smoking related phenotypes. Among these, the most strongly associated single nucleotide polymorphism (SNP) in several genome-wide association studies (GWAS) of nicotine dependence and correlated traits is a common non-synonymous change (rs16969968/D398N) in CHRNA5 43. Despite big differences in the frequency of this SNP across populations it shows a similar effect on risk, with odds ratios 1.3-2.0 in European, African American and Asian populations 38,119. In vitro functional studies have demonstrated that receptors including the amino acid encoded by the minor allele, which is the risk allele, show a reduced response to the nicotine agonist, epibatidine 85. Human carriers of the minor allele show decreased intrinsic resting connectivity strength in the dorsal anterior cingulate–ventral striatum/extended amygdala circuit 89. Additionally, a group of highly correlated variants including the SNP rs588765 near CHRNA5 were shown to increase CHRNA5 mRNA expression and increase risk of nicotine dependence independently of rs16969968 41,45. Common variants at the CHRNA6-CHRN3 gene cluster were also recently shown to contribute to smoking quantity 43. Together these results strongly suggest multiple mechanisms connecting nicotinic receptor function to nicotine dependence. Despite these findings, only a small proportion of the variance (~5%) in nicotine dependence related traits have been explained by these variants 45. Neuronal nicotinic acetylcholine receptors are pentameric ion channels produced from various combinations of α and β receptor
subunits and are the main target for nicotine in the brain. The vast majority of neuronally expressed nicotinic receptors contain the α4 and β2 subunits, but many other combinations are formed and have distinct properties and expression patterns. As such, it is thought that the various aspects of nicotine dependence (craving, withdrawal, tolerance, sensitivity, etc.) may be due to these distinct characteristics. For example, recent studies have shown that mice lacking Chrna5 fail to display the adverse effects of high doses of nicotine and mice lacking Chrnb4 or Chrna5 show reduced somatic signs of withdrawal after ceasing long-term nicotine treatment. Additionally, mice over-expressing Chrnb4 were recently shown to display a strong aversion to nicotine, that can be reversed by virally mediated expression of α5 D398N variant in the medial habenula. This work suggests that aversion of nicotine in mouse is regulated by the balanced activity of β4 and α5 nicotinic receptor subunits in the medial habenula. A recent association study of smoking cessation also suggests that variants in CHRNA5 may decrease craving and withdrawal symptoms. These findings suggest that multiple nicotinic receptor subunit genes likely play a role in the development and maintenance of nicotine dependence and that variants in these genes may have different and in some cases opposing functional effects.

Rare genetic variants have recently been shown to contribute to a number of common human diseases (see for review). Specifically, multiple rare variants have been reported in genes previously shown to harbor common variants associated with common diseases. Additionally, studies have demonstrated associations between rare variation in nicotinic receptor genes and nicotine dependence. As a class, rare SNPs, copy-number variants (CNVs) and small insertion/deletion polymorphisms (indels) constitute
the majority of human genetic variation and thus may hold the key to understanding part
of the missing heritability of complex traits unaccounted for by recent genome-wide
association studies. Sequencing is one of the few methods by which these variants can be
investigated; however, the cost of sequencing hundreds or thousands of individuals for
multiple genes remains high despite advances in sequencing technologies. An efficient
alternative approach is to carry out DNA sequencing using DNA pooled from multiple
individuals. To test the hypothesis that rare variants in the CHRNA5, CHRNA3, CHRNB4,
CHRNA6 and CHRNB3 genes influence risk for nicotine dependence we have employed
pooled sequencing in individuals with and without nicotine dependence in both European
Americans and African Americans. We report that rare (<5%) missense variants at
conserved sites in CHRNB4 are associated with decreased risk of developing nicotine
dependence and decrease the number of cigarettes consumed daily by smokers. These
findings demonstrate that nicotine dependence is mediated by rare genetic variation and
by variation in CHRNB4 specifically.

METHODS

Sample Selection

DNA samples were collected as part of the Collaborative Genetic Study of Nicotine
Dependence (COGEND). All members of the COGEND sample underwent a semi-
structured interview, which assessed smoking behavior, other substance use and
comorbid psychiatric conditions. The COGEND sample includes 710 African Americans
(461 nicotine dependent (ND) cases and 249 smokers with no symptoms of dependence
(controls)) and 2055 European Americans (1062 ND cases and 993 controls). As these
individuals were ascertained to study nicotine dependence, Fagerström Test of Nicotine Dependence (FTND) scores (Range= 1-10) were required to be \( \leq 1 \) for controls and \( \geq 4 \) for cases. In all cases lifetime maximum FTND score was used. Additionally, to ensure that all individuals in the study had been exposed to nicotine, all members were required to have smoked at least 100 cigarettes in their lifetime. Among individuals with FTND scores \( \geq 4 \), 96% (1458/1523) were also nicotine dependent using DSM-IV nicotine dependence criteria. A total of 352 African Americans (176 ND cases and 176 smoking controls) and 400 European Americans (200 ND cases and 200 smoking controls) were sequenced. For African Americans, pools of 88 individuals were used (2 case pools and 2 control pools). For European Americans, pools of 200 were used (1 case pool and 1 control pool). The case pools include individuals with FTND scores \( \geq 5 \) whereas the control pools include individuals with FTND scores \( \leq 1 \). Individuals for case pools were chosen to select the most severely nicotine dependent subjects in our sample. Follow up genotyping of SNPs identified and validated in the sequenced individuals was done in the remaining portion of the COGEND sample (310 African Americans and 1662 European Americans).

**Pooled Sequencing**

Pooled DNA sequencing was performed as previously described \(^{120}\). A schematic of the technique can be seen in figure 2.1. The concentrations of individual DNA samples were first measured using Quant-iTTM PicoGreen reagent and pooled in equimolar amounts. Each pooled DNA sample was then used as the template for the amplification of each protein coding exon of the \( {CHR}A3 \), \( {CHR}A5 \), \( {CHR}A6 \), \( {CHR}B3 \) and \( {CHR}B4 \) genes. Primers for the amplification of each exon were designed using Primer3 and reference
sequences were taken from the human genome reference assembly build 36.1 (hg18). In order to ensure complete coverage of each desired exon, a minimum of 50 bp of flanking sequence on each side was required for each amplicon. We used Pfu high fidelity DNA polymerase in all PCR reactions to reduce the identification of SNPs generated as a result of the PCR (false positive SNPs). After PCR amplification of desired genomic regions, PCR products were cleaned using QIAquick PCR purification kits, quantified using Quant-iT PicoGreen reagent and ligated in equimolar amounts using T4 Ligase and T4 Polynucleotide Kinase. At this stage, positive and negative control vectors were amplified and added to each pool to serve as internal quality standards and to be used in data analysis. After ligation, concatenated PCR products were randomly sheared by sonication and prepared for Illumina sequencing on an Illumina Genome Analyzer IIx (GAIIx) according to the manufacturers specifications. As previously shown by Vallania et al., an average coverage of 30-fold per allele per pool was shown to correlate to optimal positive predictive value for the SNP-calling algorithm and was, therefore, the target level of coverage in this study. In order to obtain sufficient coverage per allele per pool, we obtained two lanes of Illumina GAIIx sequencing per pool. Coverage per amplicon was calculated for each pool after the first lane of sequence was obtained and only those amplicons not reaching 30-fold coverage per allele were included for the second lane of sequencing.

**Sequencing Analysis**

For analysis, sequencing reads (36 bp reads) were aligned using an alignment algorithm developed by Vallania et al. which aligns sequences allowing for 2 mismatches or indels of up to 4 bp. In order to quantify the specificity and sensitivity of this method,
positive and negative control DNA were introduced as PCR products in the pooled sequencing protocol outlined above. As a positive control to estimate sensitivity for variant calling, a pool of 10 plasmids with a 72 bp insert was generated. One plasmid acts as the “wild-type” insert, while the remaining nine plasmids contain one or two synthetically engineered mutations. All ten plasmids were combined such that the allele frequency of each known mutation would mimic either the allele frequency of a single allelic variant or ten allelic variants in our human DNA pool. Following mixing of the vectors, PCR amplification across the insert sequence was performed and the PCR product was added to the normalized pool of human target PCR reactions during sequencing library preparation. We then obtained more coverage for each amplicon than that which was required to detect all variants in this amplified positive control vector pool in order to ensure a minimal SNP detection false negative rate. As a negative control and to model the sequencing error rate, 1908 bp of the pCMV6-XL5 plasmid were also amplified and included. One half of the pCMV6-XL5 plasmid sequence was used to train the SNP finding algorithm and the other to test the error model. Finally, to identify variants, in the pooled sequence data we used the SPLINTER algorithm. The SPLINTER algorithm produces a p-value for the probability that a predicted variant is a true positive. This probability is produced using large deviation theory to quantify the difference between observed allele frequencies within the sequence data to the background error rate for the same type of sequence changes seen in the amplified pCMV6-XL5 plasmid. The error model for one of the sequenced COGEND pools can been seen in figure 2.2. A p-value cut-off value for each lane of Illumina sequencing was defined as the value at which all positive controls were identified. Only those variants
falling below this cut-off value were considered “predicted” by SPLINTER. All protein coding or splice site variants predicted by SPLINTER were then validated by individual Sequenom genotyping in each person from the source DNA pools. SNPs validated in the sequenced individuals were then genotyped in all members of the COGEND sample. All individual genotyping was performed using the Sequenom platform as described previously \(^{41}\).

**Association Analysis**

Each missense variant as well as the number of carriers of missense variants at conserved positions for each gene in the genes surveyed was tested for association with nicotine dependence in African Americans and European Americans using Fisher’s exact test. Conservation was determined by basewise vertebrate conservation using PhyloP score \(^{122}\). A site was called conserved when its PhyloP score was greater than or equal to 2, corresponding to a p-value of 0.01. The SNP rs56218866 (*CHRNB4* S140G) was excluded from analyses as it had been shown previously to have no measurable effect on receptor function and did not effect the significance of the association between *CHRNB4* carrier status and either nicotine dependence or CPD when taken as a covariate \(^{88}\). For analysis of cigarettes per day (CPD), FTND score and withdrawal symptom count in carriers and non-carriers, Wilcoxon Rank-Sum tests were performed.

**RESULTS**

In order to identify novel rare variants associated with nicotine dependence, we sequenced 752 individuals from the extremes of the population distribution of FTND
scores in 6 pools. The protein coding regions of the 5 human cholinergic nicotinic receptor subunit genes constituting the gene clusters previously reported to harbor common variants associated with nicotine dependence (CHRNA5-CHRNA3-CHRNB4 and CHRNA6-CHRNB3) were sequenced on an Illumina GAIIx and analyzed using the SPLINTER algorithm. We obtained greater than 30-fold coverage per allele at all positions within the 28 amplicons designed to cover the protein coding exons of CHRNA3, CHRNA5, CHRNA6, CHRNB3 and CHRNB4 and validated 24 (9 novel) missense variants (Figure 2.3 and Table 2.1). We only attempted to validate missense variants as these are more likely to have a functional impact than non-coding or synonymous coding changes. As the SPLINTER algorithm produces a p-value for the probability that a predicted variant is a true positive, we also genotyped a number of variants below our cut-off for significance in order to determine the positive predictive value of the algorithm. A total of 42 variants were genotyped in the sequenced individuals. Of these, 14 were selected for having probabilities of being true positives just below our cut-off value. Only 1 of these 14 variants was validated while 24 of 28 SNPs predicted to be true positives were validated. This produced a positive predictive value of 86%. Additionally, all variants included as part of the amplified positive control vector were found upon achieving greater than 30-fold coverage at mutated sites (Sensitivity = 100%) and only ~80 sites in the 1908 bp negative control vector were predicted to be polymorphic (Specificity = ~95%). The majority of validated variants, as expected, are rare (22/24 = 92% have MAF < 5%) in African Americans or Europeans. In fact, many are present in only one sequenced individual (7/24 = 29% are singletons). Additionally, some of the missense variants present in these genes were found at less conserved sites.
(9/25; PhyloP score < 2), suggesting that they are less likely to have a functional impact. Altogether, we identified 15 variants (4 novel) at conserved sites and used these for further analysis. Of these, 12 were also predicted to be damaging by Polyphen, SIFT or both (Table 2.2)123,124. We determined the frequency of each missense variant validated from sequencing in individuals from the Collaborative Study on the Genetics of Nicotine Dependence (COGEND) (710 African Americans and 2055 European Americans) using Sequenom. We then compared the frequency of each missense SNP as well as the number of carriers of conserved missense variants in cases and controls for each of the genes sequenced (Table 2.2). To ensure novel findings, we excluded the SNP rs16969968 from analyses in European Americans as this would overshadow the effects of other variants in the *CHRNA5* gene and because it is common in European Americans (Frequency= 0.35). We find that two SNPs in *CHRNA4* (T91I and T375I) and one SNP in *CHRNB3* (R37H) are associated with decreased risk of nicotine dependence either in African Americans or European Americans (Table 2.1). Notably, the T91I and R37H variants are in high linkage disequilibrium (LD) in European Americans (r^2 =0.89, n=710) and African Americans (r^2 =0.59, n=2035), making it difficult to determine the relative contribution of the two variants to the observed association. We found nominal evidence of association between rare variants at evolutionarily conserved sites and nicotine dependence at *CHRNA5* in African Americans and *CHRNB3* in European Americans.

We find that carriers of rare missense variants at conserved sites in *CHRNA4* were less frequently nicotine dependent compared to non-carriers, both in African Americans and European Americans (African Americans p = 0.0025, OR = 0.31, 95% CI = 0.31-
0.72; European Americans p = 0.023, OR=0.69, 95% CI = 0.50-0.95) (Table 2.3). This is significant for African Americans after Bonferroni correction for multiple tests assuming one test for each gene studied (p < 0.01). As we are not able to distinguish between the effects of the T91I and R37H variants, carriers of missense variants at conserved sites in either CHRNA3 or CHRN4 are also less frequently nicotine dependent compared to non-carriers (African American p = 0.0035 European American p = 0.012). To determine if the observed association could be explained by the CHRNA3 R37H alone, we performed logistic regression including CHRNA3 R37H genotype as a covariate. CHRN4 carrier status remained associated with nicotine dependence status in African Americans (p = 0.0025), but not in European Americans (p=0.37). This is most likely because rs61737499 (CHRNB4 T375I), one of the major contributors to the association SNP in African Americans, is monomorphic in European Americans. Additionally, we investigated the possibility that one of the four comorbid DSM-IV defined substance dependence phenotypes measured in our dataset (alcohol, cocaine, marijuana and opiates) was contributing to the observed association at CHRN4. None of these comorbid phenotypes significantly altered the association results observed between nicotine dependence and CHRN4 carrier status in either African Americans or European Americans when included as a covariate in a logistic regression. These findings suggest that variants in CHRN4 and possibly CHRNA3 protect against nicotine dependence and that this is not due to a correlation between nicotine dependence and other comorbid phenotypes.

Two groups of common genetic variants within the CHRNA5-CHRNA3-CHRN4 gene cluster (tagged by rs16969968 and rs588765, respectively) were previously shown
to affect nicotine dependence risk. To determine if our findings could be due to LD with either of these variants, we performed logistic regression including each as a covariate. 

*CHRNB4* carrier status remained associated with nicotine dependence status in African Americans (p = 0.0048), but not in European Americans (p=0.13) when rs16969968 genotype was included as a covariate. However, we find that *CHRNB4* carrier status is associated with nicotine dependence among carriers of the minor allele of rs16969968 (Fisher’s exact test p=0.03, OR=0.55) in European Americans. *CHRNB4* carrier status was associated with nicotine dependence status in African Americans (p = 0.0031) and in European Americans (p=0.043) when rs588765 genotype was included as a covariate. These findings suggest that our observed associations are not due to a correlation between rare variants in *CHRNB4* and previously described common variant associations in the region.

To ensure that our findings within the African American portion of the COGEND sample is not due to population stratification, we calculated principal components using EIGENSTRAT. We then performed logistic regression to test the association between *CHRNB4* carrier status at conserved sites and nicotine dependence including the first two principal components (PC1 and PC2) as covariates. *CHRNB4* carrier status remained associated with nicotine dependence status in African Americans (p = 0.0038).

Additionally, we calculated local admixture using LAMP. *CHRNB4* carrier status remained significantly associated with nicotine dependence when we performed logistic regression including local admixture estimates from the 100kb encompassing the *CHRNA5-CHRNA3-CHRNB4* gene cluster as a covariate (p = 0.0029). These findings
suggest that the observed association is not due to population stratification, either on a genome scale or on a local scale.

In order to examine which aspects of nicotine dependence are affected by variants in \textit{CHRN}B4, we tested the association between several nicotine related phenotypes in addition to case status. When we compared the distribution of CPD in \textit{CHRN}B4 conserved SNP carriers and non-carriers, we observed that carriers also have a lower mean CPD than non-carriers (African American \(p = 6.6 \times 10^{-5}\), European American \(p = 0.021\)) (Figure 2.4). FTND as a quantitative trait was also lower in carriers compared to non-carriers (African American \(p = 0.0019\), European American \(p = 0.008\)) (Table 2.3). Using DSM-IV criteria to assign nicotine dependence case status we find carriers to have reduced risk of becoming nicotine dependent in African Americans \(p = 0.03\) and European Americans \(p = 0.02\). As \textit{Chrnb}4 knockout mice exhibit reduced withdrawal symptoms upon cessation of prolonged nicotine exposure\(^{90,127}\), we investigated whether conserved missense variants in \textit{CHRN}B4 contribute to withdrawal symptoms among individuals who attempt cessation. We find carriers of conserved missense variants in \textit{CHRN}B4 show fewer signs of withdrawal as measured by the number of withdrawal symptoms endorsed in African Americans but not European Americans (African American \(p = 0.012\), European American \(p = 0.42\)) (Table 2.3).

\textbf{DISCUSSION}

We find that rare missense variants at evolutionarily conserved sites in \textit{CHRN}B4 and \textit{CHR}NA3 protect against nicotine dependence in African Americans and European
Americans. We do not find an association between rare variants, conserved or otherwise, and nicotine dependence at the other genes studied (CHRNB3, CHRNA6 and CHRNA5) in African Americans or European Americans. This may be due to insufficient power resulting from low minor allele frequencies even after collapsing genotypes within genes. For instance, the frequency of carriers of missense variants at conserved loci for CHRNA5 in European Americans is 0.015 while the frequency of such carriers for CHRNA5 in European Americans is 0.055. This increased carrier frequency for CHRNA5 was in large part due to the combined frequency of the two variants (T375I and T911) which is why we chose to study these SNPs in vitro (see chapter 4). It is important to note that rare variant associations are more dependent on stochastic events, namely the sampling of rare allele carriers from a population, than studies of common genetic variants. It is possible that our observed associations may be due to chance fluctuations in rare variant allele.

When considering genetic associations in admixed populations, it is important to insure that findings are not due to subtle population stratification, either on a genome scale or locally at associated loci. As a result, we calculated both global and local measures of admixture for each of the African American individuals in COGEND. We find that neither global nor local admixture is able to account for the association we observe between rare missense variants at evolutionarily conserved sites in CHRNA5 and nicotine dependence.

In addition to nicotine dependence, we examined the effect of conserved missense variants in CHRNA5 on several phenotypes designed to tease apart various aspects of dependence. Specifically, we find that African American and European American rare variant carriers smoke fewer cigarettes per day (CPD) and that African American carriers
experience fewer withdrawal symptoms when compared to non-carriers. These data suggest that these CHRNB4 variants are likely contributing to the risk of developing nicotine dependence by allowing carriers to quit more easily, possibly by allowing them to smoke fewer cigarettes per day to combat withdrawal symptoms. It is difficult to determine with confidence which aspect of nicotine dependence variants in CHRNB4 are affecting, however, as each of the phenotypes analyzed are highly correlated in our sample ($r^2$ from 0.32-0.85) and because our sample was selected to contain the extremes of the population distribution of FTND score. These findings are in contrast to those of Wessel et al. who found no association between CHRNB4 variants and nicotine dependence. This lack of association on their part, however, was likely due to lack of power (they sequenced on 448 individuals) and our association is strongest in African Americans, a population not sequenced at these genes previously by any group.

In conclusion, we find an excess of conserved missense variants in CHRNB4 in smokers who are not nicotine dependent compared to smokers who are nicotine dependent. We hypothesize that these variants act by increasing sensitivity to nicotine and that these changes lead to our overall finding of decreased risk for nicotine dependence. Further sequencing of individuals for whom detailed dependence related information has been obtained is necessary to confirm and elaborate on these findings.
Figure 2.1. Pooled Sequencing Scheme for COGEND. DNA was quantified and pooled. The exons and flanking regions for the listed genes were amplified by PCR using PFU high-fidelity polymerase, pooled and sequenced on an Illumina GAIIx.
Figure 2.2. Representative Error Model for COGEND Pools. Each line is the average rate of change of each nucleotide changing to each of the other possible nucleotides for each of the possible positions within the Illumina sequencing read. Error rates increase dramatically after 30 cycles as a result of the chemistry used.
Figure 2.3. Schematics of the CHRN genes studied and the locations of missense variants identified in sequenced individuals from the COGEND study. Each protein consists of a signal peptide, two extracellular domains, four transmembrane domains (TM1–4) and three intracellular domains. Conserved sites (phyloP44Vertebrate score ≥ 2) are shown in grey, while all others are shown in black.
<table>
<thead>
<tr>
<th>Gene</th>
<th>AA Change</th>
<th>RS Number</th>
<th>CHR</th>
<th>Position</th>
<th>A1</th>
<th>A2</th>
<th>PhyloP Score</th>
<th>SIFT</th>
<th>Polyphen</th>
<th>African Americans</th>
<th>European Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>S326T</td>
<td>8</td>
<td>42587426</td>
<td>T</td>
<td>A</td>
<td>0.67</td>
<td></td>
<td>0.002</td>
<td>0.002</td>
<td>1</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>H329Y</td>
<td>8</td>
<td>42587435</td>
<td>C</td>
<td>T</td>
<td>3.98</td>
<td>+</td>
<td>0.058</td>
<td>0.045</td>
<td>1.30</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>K451E</td>
<td>rs35327613</td>
<td>8</td>
<td>42591735</td>
<td>G</td>
<td>A</td>
<td>2.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>H455N</td>
<td>-</td>
<td>8</td>
<td>42591747</td>
<td>C</td>
<td>A</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>N447S</td>
<td>rs16891583</td>
<td>8</td>
<td>42611002</td>
<td>C</td>
<td>T</td>
<td>4.96</td>
<td>++</td>
<td>0.001</td>
<td>0</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>T231I</td>
<td>-</td>
<td>8</td>
<td>42611650</td>
<td>G</td>
<td>A</td>
<td>6.91</td>
<td>++</td>
<td>0.002</td>
<td>0.46</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>R96H</td>
<td>-</td>
<td>8</td>
<td>42621158</td>
<td>C</td>
<td>T</td>
<td>0.94</td>
<td>+</td>
<td>0.001</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>S43P</td>
<td>-</td>
<td>8</td>
<td>42620300</td>
<td>A</td>
<td>G</td>
<td>1.23</td>
<td>+</td>
<td>0.011</td>
<td>0.59</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>S81F</td>
<td>-</td>
<td>15</td>
<td>78873288</td>
<td>C</td>
<td>T</td>
<td>5.35</td>
<td>+</td>
<td>0.012</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>V134I</td>
<td>rs2229961</td>
<td>15</td>
<td>78806752</td>
<td>A</td>
<td>G</td>
<td>6.72</td>
<td>++</td>
<td>0.012</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>K167R</td>
<td>-</td>
<td>15</td>
<td>78882233</td>
<td>A</td>
<td>G</td>
<td>5.40</td>
<td>++</td>
<td>0.012</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>L363Q</td>
<td>rs79109919</td>
<td>15</td>
<td>78882821</td>
<td>T</td>
<td>A</td>
<td>4.77</td>
<td>+</td>
<td>0.061</td>
<td>0</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>D398N</td>
<td>rs16969968</td>
<td>15</td>
<td>78882925</td>
<td>A</td>
<td>G</td>
<td>3.77</td>
<td></td>
<td>0.034</td>
<td>0</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>R37H</td>
<td>rs8192475</td>
<td>15</td>
<td>78911230</td>
<td>C</td>
<td>T</td>
<td>2.67</td>
<td>++</td>
<td>0.008</td>
<td>0.75</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>M467V</td>
<td>rs51737502</td>
<td>15</td>
<td>78917735</td>
<td>T</td>
<td>C</td>
<td>1.58</td>
<td></td>
<td>0.083</td>
<td>0</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>M456V</td>
<td>-</td>
<td>15</td>
<td>78917606</td>
<td>T</td>
<td>C</td>
<td>4.71</td>
<td>++</td>
<td>0.008</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>T375I</td>
<td>rs51737499</td>
<td>15</td>
<td>78921523</td>
<td>G</td>
<td>A</td>
<td>2.26</td>
<td>+</td>
<td>0.008</td>
<td>0.20</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>R349C</td>
<td>-</td>
<td>15</td>
<td>78921602</td>
<td>A</td>
<td>G</td>
<td>1.96</td>
<td>++</td>
<td>0.001</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>G296S</td>
<td>-</td>
<td>15</td>
<td>78921761</td>
<td>C</td>
<td>T</td>
<td>5.69</td>
<td>++</td>
<td>0.012</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>S140G</td>
<td>rs56218866</td>
<td>15</td>
<td>78922229</td>
<td>C</td>
<td>T</td>
<td>2.27</td>
<td>+</td>
<td>0.047</td>
<td>1.20</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>R136Q</td>
<td>rs56095004</td>
<td>15</td>
<td>78922240</td>
<td>C</td>
<td>T</td>
<td>0.41</td>
<td>+</td>
<td>0.033</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>R136W</td>
<td>-</td>
<td>15</td>
<td>78922241</td>
<td>G</td>
<td>A</td>
<td>0.33</td>
<td>++</td>
<td>0.012</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>T91I</td>
<td>rs12914008</td>
<td>15</td>
<td>78923505</td>
<td>A</td>
<td>G</td>
<td>2.16</td>
<td></td>
<td>0.005</td>
<td>0</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>K57E</td>
<td>-</td>
<td>15</td>
<td>78927816</td>
<td>T</td>
<td>C</td>
<td>1.16</td>
<td></td>
<td>0.001</td>
<td>0</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1. Missense Variants in sequenced nicotinic receptor genes present in the COGEND sample. Cases are individuals with FTND score ≥4 and controls are individuals with FTND scores ≤1. PhyloP score is the base-wise vertebrate conservation score. Positions are from human genome reference assembly build 37 (hg19).
<table>
<thead>
<tr>
<th></th>
<th>African Americans (N=710)</th>
<th>European Americans (N=2055)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case Frequency</td>
<td>Control Frequency</td>
</tr>
<tr>
<td><strong>CHRNA5</strong></td>
<td>0.139</td>
<td>0.102</td>
</tr>
<tr>
<td><strong>CHRNA6</strong></td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>CHRNA3</strong></td>
<td>0.057</td>
<td>0.044</td>
</tr>
<tr>
<td><strong>CHRNA4</strong></td>
<td>0.012</td>
<td>0.036</td>
</tr>
<tr>
<td><strong>CHRNA5</strong></td>
<td>0.008</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Table 2.2. Gene-based Association Results. Fisher’s exact tests were used to determine the degree of association between nicotine dependence and carrying at least one missense variants at evolutionarily conserved positions (vertebrate PhyloP > 2).
### CHRNAB4

<table>
<thead>
<tr>
<th></th>
<th>African Americans</th>
<th></th>
<th></th>
<th>European Americans</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Carriers</td>
<td>Carriers</td>
<td>P-value</td>
<td>Non-Carriers</td>
<td>Carriers</td>
<td>P-value</td>
</tr>
<tr>
<td>n</td>
<td>681</td>
<td>29</td>
<td>0.64</td>
<td>1857</td>
<td>178</td>
<td>0.72</td>
</tr>
<tr>
<td>Age</td>
<td>37 ± 6</td>
<td>37 ± 7</td>
<td></td>
<td>37 ± 5</td>
<td>36 ± 6</td>
<td>0.72</td>
</tr>
<tr>
<td>Women/Men</td>
<td>250/431</td>
<td>11/18</td>
<td>1.00</td>
<td>727/1151</td>
<td>67/110</td>
<td>0.87</td>
</tr>
<tr>
<td>Nicotine Cases / Controls</td>
<td>450/231</td>
<td>11/18</td>
<td><strong>0.0025</strong></td>
<td>974/883</td>
<td>77/101</td>
<td><strong>0.023</strong></td>
</tr>
<tr>
<td>Cigarettes per Day</td>
<td>17 ± 13</td>
<td>9 ± 6</td>
<td><strong>6.6 \times 10^{-5}</strong></td>
<td>19 ± 17</td>
<td>16 ± 14</td>
<td><strong>0.021</strong></td>
</tr>
<tr>
<td>FTND Score</td>
<td>4.0 ± 3.2</td>
<td>2.1 ± 2.8</td>
<td><strong>0.0019</strong></td>
<td>3.4 ± 3.5</td>
<td>2.7 ± 3.3</td>
<td><strong>0.008</strong></td>
</tr>
<tr>
<td>Withdrawal Symptom Count</td>
<td>3.7 ± 2.4</td>
<td>2.4 ± 2.9</td>
<td><strong>0.012</strong></td>
<td>3.4 ± 2.5</td>
<td>3.2 ± 2.4</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Table 2.3. Demographic and Phenotypic Characteristics of COGEND African Americans and European Americans. Carriers are individuals who possess at least one of the following: T91I, T375I, G296S, or M456V. Nicotine dependent cases are maximum lifetime FTND ≥ 4 while controls are maximum lifetime FTND < 1. Cigarette per day (CPD) values are mean ± s.d. and represent the average number of cigarettes smoked per day when the individual was smoking the most. FTND score is maximum lifetime FTND score. P values were calculated using a two-sided Fisher’s exact test or two-sided Wilcoxon rank-sum test.
Figure 2.4. Average number of cigarettes smoked per day (CPD) for carriers and non-carriers of missense variants at evolutionarily conserved sites in African Americans and European Americans. *** = 6.6x10^{-5}; * = 0.021
ABSTRACT

Previous findings have demonstrated that variants in nicotinic receptor genes are associated with nicotine, alcohol and cocaine dependence. Because of the substantial comorbidity, it has often been unclear whether a variant is associated with multiple substances or whether the association is actually with a single substance. In order to investigate the possible contribution of rare variants to the development of substance dependencies other than nicotine dependence, specifically alcohol and cocaine dependence, we undertook pooled sequencing of the coding regions and flanking sequence of CHRNA5, CHRNA3, CHRNA6, CHRNA7, and CHRNA8 in 287 African American and 1028 European American individuals from the Collaborative Study of the Genetics of Alcoholism (COGA). All members of families for whom any individual was sequenced (2504 African Americans and 7318 European Americans) were then genotyped for all variants identified by sequencing. For each gene, we then tested for association using FamSKAT, which incorporates family structure using a kinship matrix. For European Americans, we find increased DSM-IV cocaine dependence symptoms (FamSKAT p =2x10^{-4}) and increased DSM-IV alcohol dependence symptoms (FamSKAT p=5x10^{-4}) among carriers of missense variants in CHRNA3. Additionally, one variant (rs149775276; H329Y) shows association both with cocaine dependence symptoms (p=7.4x10^{-5}, β=2.04) and alcohol dependence symptoms (p=2.6x10^{-4}, β=2.04)

3 Portions of this chapter are adapted from: Haller, G., Kapoor, M., et al. Rare missense variants in CHRNA3 and CHRNA7 are associated with risk of alcohol and cocaine dependence (submitted)
when analyzed alone. For African Americans, we find decreased cocaine dependence symptoms among carriers of missense variants in \textit{CHRNA3} (FamSKAT p=0.006). These are the first results to implicate rare variants in \textit{CHRNB3} or \textit{CHRNA3} in risk for alcohol dependence or cocaine dependence.

\textbf{INTRODUCTION}

Twin and adoption studies have suggested that in addition to genetic variants that contribute to the risk of becoming dependent on a specific drug, there exist genetic factors underlying a generalized risk for becoming dependent on multiple substances (see \textsuperscript{58} for review, \textsuperscript{59,60}). Epidemiological studies have shown high levels of comorbidity among alcohol, cocaine and nicotine dependence \textsuperscript{128,129}. Because of the substantial comorbidity, it is often unclear whether a variant is associated with multiple substances or whether the association is actually with a single substance but the extensive comorbidity leads to apparent association with other substances. Numerous studies have found that variants altering the function or expression of neuronal cholinergic nicotinic receptors (\textit{CHRN}s) alter risk for becoming nicotine dependent and several have shown effects of variants in nicotinic receptors on risk for both alcohol and cocaine dependence (see \textsuperscript{35} for review). Common single nucleotide polymorphisms (SNPs) in \textit{CHRNA5} and \textit{CHRNB3} are associated with nicotine dependence or cigarette consumption \textsuperscript{43,130}. The most strongly associated SNP in several GWAS of nicotine dependence, a non-synonymous change (rs16969968/D398N) in \textit{CHRNA5}, has also been shown to mitigate risk for cocaine dependence \textsuperscript{61,62}. Additionally, a highly correlated group of variants near \textit{CHRNA5}, previously shown to alter \textit{CHRNA5} mRNA expression \textit{in vivo}, have been reported to alter risk for alcohol dependence \textsuperscript{41} in addition to nicotine dependence \textsuperscript{131}. 

54
GWAS of cigarette consumption has also implicated a group of common SNPs near the
*CHRNB3-CHRNA6* gene cluster\(^{43}\) that have been suggested to be involved in alcohol
consumption, although not with cocaine related behaviors\(^{132}\). In addition to mediating the
effects of nicotine, neuronal nicotinic receptors have also been shown to alter the cellular
and behavioral effects of alcohol and cocaine independently of nicotine\(^{92-94,97-114}\).
Evidence from *in vitro* studies indicates altered response to ethanol upon perturbation of
nicotinic receptors in neurons\(^{95,106-108,115}\). Studies in rats have shown that partial agonists
of \(\alpha_4\beta_2^*\) or \(\alpha_3\beta_4^*\) nicotinic receptors affect alcohol consumption *in vivo*\(^{92,95-98,101-105}\).
These results were confirmed and expanded using transgenic and knockout mice for
various nicotinic receptor genes\(^{93,94,109-114}\). The physiological and behavioral effects of
cocaine are also dependent on nicotinic receptors. Studies in mice, rats and non-human
primates have suggested that cocaine administration alters nicotinic receptor expression
and in so doing leads to altered response to cocaine\(^{133-136}\). A recent human trial also
found that the \(\alpha_4\beta_2\) subunit partial agonist Varenicline reduced alcohol consumption
among smokers\(^{116}\). Together these results suggest that altered nicotinic receptor
expression and function likely has an effect on alcohol and cocaine consumption
independent of its effect on nicotine consumption. However, despite the promising
evidence of synergy and cross-substance influences, the combined effects of discovered
common variants have explained only a small proportion of the variance (~5\%) in
nicotine dependence and even less of the variance in alcohol or cocaine dependence\(^{45}\).
Rare variants may explain much of the unexplained heritability of these complex
substance use disorders.
As a class, rare variants constitute the majority of genetic variation. A recent survey of more than 4000 exomes from individuals of European and African descent showed that as much as 86% of coding variants have a frequency less than 0.05% and that these rare variants are 4 times as likely to be deleterious. Many examples now exist of rare variants contributing to common human diseases (see for review), both in genes previously reported to cause Mendelian forms of disease and in genes harboring common SNP associations. Several studies have demonstrated associations between rare variation in nicotinic receptor genes and nicotine dependence. An efficient method of determining the full spectrum of genetic variation in a region and thus the relative contribution of rare genetic variation to a trait of interest is to sequence pools of DNA from multiple individuals and to use statistical methods to determine which sites are polymorphic and at what frequency. We have employed this approach to determine the contribution of rare non-synonymous variants in the CHRNA5, CHRNA3, CHRNA4, CHRNA6 and CHRNA3 genes to alcohol and cocaine dependence. My aim in this chapter is to investigate the hypothesis that rare missense variants in nicotinic acetylcholine receptor genes contribute in aggregate to risk for alcohol and/or cocaine dependence independently of nicotine dependence.

METHODS

Sample Selection

COGA Sample
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 2504 African Americans and 7318 European Americans. Maximum lifetime cigarettes per day (CPD) was used as a covariate in all analyses. CPD was calculated by comparing reports across multiple interviews for an individual and selecting the value of cigarettes smoked per day that reflected the largest number smoked during the time when the subject was smoking the most. Only individuals who had at one point in their life smoked have CPD values. Controls consisted of individuals who had tried a substance but never became dependent. A total of 287 African American (147 cases and 140 controls) and 1028 European American (480 cases and 548 controls) individuals were sequenced. For African Americans, pools ranging in size from 44-96 individuals were used (2 case pools and 2 control pools). For European Americans, pools ranging in size from 68-96 individuals were used (5 case pools and 6 control pools). Follow up genotyping of SNPs identified and validated in the sequenced individuals was done in the remaining members of each family represented in the sequenced set of individuals (2217 African Americans and 6290 European Americans).

**COGEND Sample**

DNA samples were collected as part of the Collaborative Genetic Study of Nicotine Dependence (COGEND). All members of the COGEND sample underwent a semi-structured interview, which assessed smoking behavior, other substance use and comorbid psychiatric conditions. The COGEND sample includes 710 African Americans and 2055 European Americans. The coding regions of the \( CHRNA3, CHRN4B, CHRNA5, \)
*CHRNB3* and *CHRNA6* genes were sequenced in a total of 352 African Americans (176 ND cases and 176 smoking controls) and 400 European Americans (200 ND cases and 200 smoking controls). Of these sequenced individuals, 120 (34%) of the African Americans had non-zero DSM-IV alcohol dependence symptom counts, 300 (75%) of the European Americans had non-zero DSM-IV alcohol dependence symptom counts, 23 (6.5%) of the African Americans had non-zero DSM-IV cocaine dependence symptom counts and 50 (12.5%) of the European Americans had non-zero DSM-IV cocaine dependence symptom counts. All missense variants observed in these sequenced genes were genotyped in all members of COGEND using Sequenom.

**Pooled Sequencing**

Pooled DNA sequencing was performed as previously described\(^\text{120,121,138}\). A schematic of the sequencing technique used can be seen in figure 3.1. A representative error model for one of the lanes of sequencing produced for this study can be seen in figure 3.2. As previously shown by Vallania et al,\(^\text{121}\) an average coverage of 30-fold per allele per pool was shown to correlate to optimal positive predictive value for the SNP-calling algorithm and was, therefore, the target level of coverage in this study.

**Sequencing Analysis**

For analysis, sequencing reads (36bp reads) were aligned using an alignment algorithm developed by Vallania et al.\(^\text{121}\) which aligns sequences allowing for 2 mismatches or indels of up to 4 bp. Positive and negative control DNAs were introduced as previously described\(^\text{138}\). To identify variants in the pooled sequence data we used the SPLINTER
algorithm\textsuperscript{121}. All protein coding or splice site variants predicted by SPLINTER were validated by individual Sequenom genotyping in each person from the source DNA pools. SNPs validated in the sequenced individuals were then genotyped in all members of the COGA sample. All individual genotyping was performed using the Sequenom platform as described previously\textsuperscript{139}.

**Association Analysis**

Single missense variants were tested for association in African Americans and European Americans from the COGA study using linear regression as implemented in the GWAF package in R using age, sex and CPD as covariates. Each sequenced gene was tested as a whole for association using famSKAT using age, sex and CPD as covariates. The GWAF and famSKAT programs were used as the COGA study is family-based and requires programs that can integrate family information. Associations in COGEND between alcohol dependence symptom count or cocaine dependence symptom count and variants in the \textit{CHRNB3} or \textit{CHRNA3} genes were performed using linear regression in PLINK using age, sex and CPD as covariates and in R comparing the residuals of alcohol or cocaine dependence symptom count after correcting for age, sex and CPD in \textit{CHRNB3} missense variant carriers vs. non-carriers and \textit{CHRNA3} missense variant carriers vs. non-carriers using a Wilcoxon Rank Sum Test. Linear regression and Wilcoxon Rank-Sum tests were used as COGEND is a sample of unrelated cases and controls.

**RESULTS**
We sequenced the protein coding regions of the 5 human cholinergic nicotinic receptor subunit genes constituting the gene clusters previously reported to harbor common variants associated with nicotine dependence (CHRNA5-CHRNA3-CHRNAB4 and CHRNA6-CHRNAB3) in 287 unrelated African Americans (147 DSM-IV alcohol dependent cases and 140 controls) and 1028 unrelated European Americans (480 DSM-IV alcohol dependent cases and 548 controls) as part of the Collaborative Study of the Genetics of Alcoholism (COGA). DNA pools were sequenced on an Illumina GAIIx and analyzed using the SPLINTER algorithm. In total, we validated 31 non-synonymous variants (Figure 3.3 and Table 3.1). Of these, all but rs16969968 were rare (<5% MAF), 9 (29%) were novel and 15 (48%) were present in only one sequenced individual. All variants included as part of the amplified positive control vector were found upon achieving greater than 30-fold coverage at mutated sites (Sensitivity = 100%) and only 63 sites in the 1908 bp negative control vector were predicted to be polymorphic (Specificity = 97%).

We genotyped each missense variant validated from sequencing in every individual from the COGA study for whom DNA existed, totaling 2504 African Americans and 7318 European Americans. We then tested for association between non-synonymous variant carrier status at each sequenced gene and either alcohol or cocaine dependence symptom count using the family-based Sequence Kernel Association Test (FamSKAT), which incorporates relatedness into the model using a kinship matrix. Age, sex and the number of cigarettes smoked per day (CPD) were included in the model as covariates. To increase the chances of novel findings, we did not include the SNP rs16969968 in analyses of European Americans or African Americans as it might
overshadow the effects of other variants in the \textit{CHRNA5} gene and because it is common in European Americans (Frequency= 0.35).

\textit{Analysis of European Americans}

For European Americans, we find increased DSM-IV alcohol dependence symptoms (1.3 $\pm$ 2.5 vs. 0.65 $\pm$ 1.8, famSKAT $p= 5 \times 10^{-4}$) and increased DSM-IV cocaine dependence symptoms (2.27 $\pm$ 2.7 vs. 2.07 $\pm$ 2.4, famSKAT $p = 2 \times 10^{-4}$) among carriers of missense variants in \textit{CHRNB3} (Table 3.2 and Figure 3.4). Further, among the SNPs comprising the \textit{CHRNB3} carrier genotype (S326T, H329Y, K451E and V368M), one variant (rs149775276; H329Y) was significantly associated in a linear regression model with increases in both cocaine dependence symptoms ($p= 7.4 \times 10^{-5}$, $\beta=2.04$) and alcohol dependence symptoms ($p= 2.6 \times 10^{-4}$, $\beta=2.04$) (Figure 3.4). We also tested each gene for association with CPD using famSKAT and found no significant associations between CPD and rare missense variants at any of the genes sequenced in this dataset, suggesting that this is a direct effect of these variants on risk for alcohol and cocaine dependence phenotypes.

In order to determine if these findings were the result of a correlation between alcohol or cocaine symptom count and smoking (Figure 3.5 and Figure 3.7), which had not been accounted for using CPD as a covariate, we tested whether these associations remained when only heavy smokers (CPD $\geq$20) were analyzed, as there is no significant correlation between alcohol symptom count or cocaine symptom count and CPD among these heavy smoking individuals. Using only heavy smokers (N=2509), the association between \textit{CHRNB3} and alcohol symptom count (famSKAT $p = 0.0064$) and cocaine dependence symptom count (famSKAT $p = 0.002$) remained significant, although
reduced. No associations between carrier status at any sequenced gene and alcohol or cocaine symptom count were observed in non-smokers (N=3177). Despite large numbers of non-smokers, very few non-smokers are alcohol dependent (N=387; 12%) or cocaine dependent (N=73; 2%), greatly reducing power to detect associations with these phenotypes in this subset of individuals. Overall, these analyses suggest that rare CHRNB3 variants are associated with cocaine dependence symptoms and alcohol dependence symptoms and that these associations cannot be fully explained by comorbid smoking behavior.

As expected, there are high rates of comorbidity between cocaine dependence and alcohol dependence in our sample (Figure 3.6). For example, only 3% of European cocaine dependent cases and ~5% of African Americans cocaine dependent cases in our sample have no symptoms of alcohol dependence. To determine if the observed association in European Americans between CHRNB3 variants and either cocaine dependence symptoms or alcohol dependence symptoms could be explained by a correlation between the two phenotypes, we included alcohol dependence symptom count as a covariate in the analysis of cocaine dependence and vice versa. The association between cocaine dependence symptoms and both CHRNB3 (famSKAT p=0.009) and rs149775276 (linear regression p =0.002) remained significant when alcohol dependence was used as a covariate. Using cocaine dependence symptom count as a covariate, however, CHRNB3 carrier status did not remain significantly associated with alcohol dependence symptoms (famSKAT p = 0.66). These findings suggest that variants in CHRNB3 increase risk for cocaine dependence, but only increase risk for alcohol dependence indirectly by increasing risk for cocaine dependence symptoms or that the
structure of the dataset is such that one cannot correct for cocaine dependence symptoms without disrupting all associations with alcohol dependence symptoms. One group of common SNPs within the \textit{CHRNB3-CHRNA6} gene cluster (tagged in our dataset by rs6474412) was previously shown to affect levels of nicotine consumption and nicotine dependence \textsuperscript{36,43,140}. To determine if our findings in \textit{CHRNB3} could be due to linkage disequilibrium (LD) with this variant, we tested for association between rs6474412 and both alcohol and cocaine dependence symptom count, but in neither case was a significant association observed. (alcohol symptom count \( p = 0.49, \beta = -0.03 \); cocaine symptom count: \( p = 0.91, \beta = -0.01 \)). This suggests that our observed associations are not due to a correlation between rare variants in \textit{CHRNB3} and the previously described common variant.

\textit{Analysis of African Americans}

For African Americans, we found decreased DSM-IV cocaine dependence symptoms among carriers of missense variants in \textit{CHRNA3} (famSKAT \( p = 0.006 \)) (Table3.3). As for European Americans, age, sex and CPD were included as covariates in the analysis. As there were only two missense variants observed in \textit{CHRNA3} in African Americans, this association was due largely to the association between cocaine dependence symptoms and one variant (rs8192475; R37H). This variant was previously shown to decrease risk for nicotine dependence and increase cellular response to nicotine \textit{in vitro} \textsuperscript{138}. The association between \textit{CHRNA3} carrier status and cocaine dependence symptom count remained significant even after alcohol dependence symptoms was included as a covariate (famSKAT \( p = 0.04 \)).
To determine if our findings in African Americans could be explained by subtle population stratification, we calculated principal components using EIGENSTRAT\textsuperscript{125} and included the first two principal components (PC1 and PC2) as covariates in the linear regression. \textit{CHRNA3} carrier status remained associated with cocaine dependence symptom count in African Americans (famSKAT p= 0.01). No sequenced gene other than \textit{CHRNA3} was significant before or after addition of the first two principal components as covariates in the analysis. Additionally, we calculated local admixture using LAMP\textsuperscript{126}. \textit{CHRNA3} carrier status remained significantly associated with cocaine dependence symptoms when we performed linear regression including local admixture estimates from the 100kb encompassing the \textit{CHRNA5-CHRNA3-CHRNB4} gene cluster as a covariate (famSKAT p= 0.03). No sequenced gene other than \textit{CHRNA3} was significant before or after addition of estimated local ancestry as a covariate in the analysis. These findings suggest that the observed association is not due to population stratification, either on a genome or a local scale.

\textit{Replication in COGEND}

Previous findings from the Collaborative Genetic Study of Nicotine Dependence (COGEND) have suggested that rare variants in \textit{CHRNB4} reduce risk of nicotine dependence\textsuperscript{138}. Using previously obtained sequencing data from COGEND\textsuperscript{138}, we aimed to determine whether we could replicate the findings from COGA. We tested for association between alcohol dependence symptom count and cocaine dependence symptom count and variants in the \textit{CHRNB3} and \textit{CHRNA3} genes in the COGEND sample. We used both linear regression and a non-parametric Wilcoxon Rank Sum test as appropriate for this sample of unrelated subjects. Linear regression using age, sex and
CPD as covariates revealed that, among African Americans, variants in \textit{CHRNB3} are associated with increased alcohol dependence symptoms (p= 0.004, $\beta$=0.48) (Table 3.5). This was then confirmed using a Wilcoxon Rank Sum test comparing the distribution of age, sex and CPD corrected alcohol symptom count in \textit{CHRNB3} carriers vs. non-carriers (p= 0.006). There was no association between either alcohol or cocaine symptom count and \textit{CHRNA3} variants among African Americans. More than 60\% (N=1685) of individuals in the COGEND sample have $\geq$ 1 alcohol dependence symptom, but there is a severe lack of power to interrogate cocaine dependence symptoms in COGEND because only $\sim$10\% (N=290) of individuals in this dataset have non-zero cocaine dependence symptoms. We observed no association between either \textit{CHRNB3} or \textit{CHRNA3} and cocaine dependence symptoms in European Americans from the COGEND study (Table 3.4).

**DISCUSSION**

We find that rare missense variants in \textit{CHRNB3} are associated with increased risk for cocaine dependence and alcohol dependence among European Americans and that rare missense variants in \textit{CHRNA3} are associated with decreased risk of cocaine dependence among African Americans. Upon including comorbid substance dependencies or proxies thereof (i.e. CPD for nicotine dependence) as covariates, we find that the observed association with cocaine dependence symptom count remains, suggesting that this association is not due to correlations between any of the other studied substances, but that the association between variants in \textit{CHRNB3} and alcohol dependence symptoms among European Americans may reflect the correlation of this trait with cocaine dependence symptoms in the COGA dataset.
The association between *CHRNB3* and both alcohol and cocaine dependence was largely due to one variant (rs149775276; H329Y). This variant was rare in European Americans, but absent in African Americans, partially explaining the lack of association at this gene among African Americans in the COGA dataset. We were able to replicate the association between *CHRNB3* and alcohol dependence in the COGEND African-American sample. It is likely that we were able to observe an association between *CHRNB3* variants and alcohol dependence symptoms but not cocaine dependence symptoms in COGEND because alcohol dependence symptoms appear more often among COGEND individuals (~60%) than do cocaine dependence symptoms (~10%). Additionally, this association was only observed among African Americans individuals in COGEND. The frequency of *CHRNB3* carriers among COGEND European Americans was exceedingly small (0.6% of individuals) compared to *CHRNB3* carriers among African Americans (~4%).

The association between *CHRNA3* and cocaine dependence among African Americans was due almost entirely to the effect of one variant (R37H; rs8192475). This variant was previously shown to decrease risk for nicotine dependence and increase cellular response to nicotine *in vitro*. Because *CHRNA3* R37H is rarer in African Americans (~0.1%) than European Americans (~4%), the absence of association among European Americans is not likely to be due to lack of power. However, there may be an effect of differing genetic background between the two ethnic groups that cannot be accounted for in our analyses. Additionally, controlling for both global and local admixture in the African American cohort does not affect the association observed.
between *CHRNA3* and cocaine dependence symptom counts, suggesting that this association is not due to population stratification, however subtle.

For the other genes studied (*CHRNB4*, *CHRNA6* and *CHRNA5*), no associations were observed between rare variants and either alcohol or cocaine dependence in European Americans or African Americans. This may have been due to insufficient power resulting from low minor allele frequencies even after collapsing genotypes within genes. Rare variant associations are more dependent on chance observation, namely the sampling of rare allele carriers from a population, than studies of common genetic variants. It is possible that our observed associations may be due to chance fluctuations in rare variant alleles.

There is a debate in the scientific community regarding the relative role of common versus rare genetic variants in determining an individual’s risk of developing common diseases. Whether or not the majority of variability in any given trait will eventually be attributed to common or rare variants is yet to be determined. However, this and other studies continue to demonstrate the power of rare variant association testing in the identification and validation of genes contributing to common disease. Moreover, rare variant associations have the ability to identify specific genes and variants, rather than large genomic regions. This suggests that in the era of whole genome sequencing, genome-wide gene-based rare variant associations may be more powerful in identifying novel susceptibility loci.

In conclusion, we find an association between missense variants in *CHRNB3* and *CHRNA3* and alcohol and cocaine dependence symptom counts. Variants in *CHRNB3* are associated with increased risk for cocaine dependence and alcohol dependence, while
variants in *CHRNA3* are associated with decreased risk for cocaine dependence. Both *CHRNB3* and *CHRNA3* are associated with cocaine dependence independently of smoking quantity as measured by CPD, a fact that is not surprising given the role of *CHRNB3* genes in the physiological response to multiple substances and the comorbidity of nicotine with both alcohol and cocaine use in human populations. We hypothesize that these variants act either by altering generalized reward pathways, by impacting impulsivity and related personality characteristics or by altering the effect of alcohol or cocaine on nicotinic receptors directly. Further sequencing of individuals for whom detailed dependence related information has been obtained is necessary to confirm and elaborate on these findings and functional studies will be necessary to elucidate the role of these genes on the physiological effects of cocaine.
Figure 3.1. Pooled Sequencing Scheme for COGA. DNA was quantified and pooled. The exons and flanking regions for the listed genes were amplified by PCR using PFU high-fidelity polymerase, pooled and sequenced on an Illumina GAIIx.
Figure 3.2. Representative Error Model for COGA Pools. Each line is the average rate of change of each nucleotide changing to each of the other possible nucleotides for each of the possible positions within the Illumina sequencing read.
Figure 3.3. Schematics of the CHRN genes studied and the locations of missense variants identified in sequenced individuals from the COGA study. Each protein consists of a signal peptide, two extracellular domains, four transmembrane domains (TM1-4) and three intracellular domains. Conserved sites (phyloP44Vertebrate score > 2) are shown in grey, while all others are shown in black.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Amino Acid Position</th>
<th>rs number</th>
<th>Chromosome</th>
<th>hgl9 Position</th>
<th>Allele 1</th>
<th>Allele 2</th>
<th>PhyloP Score</th>
<th>SIFT Prediction</th>
<th>Polyphen Prediction</th>
<th>Frequency in EAs</th>
<th>Frequency in AAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHRN1</td>
<td>L250P</td>
<td>-</td>
<td>8</td>
<td>4258199</td>
<td>C</td>
<td>T</td>
<td>5.117</td>
<td>+</td>
<td>++</td>
<td>0.0021</td>
<td>0</td>
</tr>
<tr>
<td>CHRN3</td>
<td>S326T</td>
<td>-</td>
<td>8</td>
<td>42587426</td>
<td>T</td>
<td>A</td>
<td>0.672</td>
<td></td>
<td></td>
<td>0</td>
<td>0.0028</td>
</tr>
<tr>
<td>CHRN3</td>
<td>H329Y</td>
<td>rs149775276</td>
<td>8</td>
<td>42587435</td>
<td>C</td>
<td>T</td>
<td>3.982</td>
<td>+</td>
<td>++</td>
<td>0.0192</td>
<td>0.0120</td>
</tr>
<tr>
<td>CHRN3</td>
<td>V368M</td>
<td>rs56198260</td>
<td>8</td>
<td>42587552</td>
<td>A</td>
<td>G</td>
<td>4.767</td>
<td></td>
<td></td>
<td>0.0188</td>
<td>0</td>
</tr>
<tr>
<td>CHRN3</td>
<td>K451E</td>
<td>rs33527613</td>
<td>8</td>
<td>42597135</td>
<td>G</td>
<td>A</td>
<td>2.350</td>
<td></td>
<td></td>
<td>0.0096</td>
<td>0.0423</td>
</tr>
<tr>
<td>CHRN6</td>
<td>N447S</td>
<td>rs16891583</td>
<td>8</td>
<td>42611002</td>
<td>T</td>
<td>C</td>
<td>3.502</td>
<td>+</td>
<td>++</td>
<td>0.0007</td>
<td>0</td>
</tr>
<tr>
<td>CHRN6</td>
<td>K408E</td>
<td>-</td>
<td>8</td>
<td>42611120</td>
<td>C</td>
<td>T</td>
<td>3.527</td>
<td></td>
<td></td>
<td>0.0028</td>
<td>0</td>
</tr>
<tr>
<td>CHRN6</td>
<td>V314G</td>
<td>rs35812517</td>
<td>8</td>
<td>42611401</td>
<td>C</td>
<td>A</td>
<td>7.129</td>
<td>+</td>
<td>++</td>
<td>0.0087</td>
<td>0.0005</td>
</tr>
<tr>
<td>CHRN6</td>
<td>L276T</td>
<td>-</td>
<td>8</td>
<td>42611665</td>
<td>C</td>
<td>T</td>
<td>5.214</td>
<td></td>
<td></td>
<td>0.0069</td>
<td>0</td>
</tr>
<tr>
<td>CHRN6</td>
<td>V110I</td>
<td>rs74376609</td>
<td>8</td>
<td>42612117</td>
<td>C</td>
<td>T</td>
<td>2.128</td>
<td></td>
<td></td>
<td>0.0035</td>
<td>0</td>
</tr>
<tr>
<td>CHRN6</td>
<td>R95H</td>
<td>-</td>
<td>8</td>
<td>42612158</td>
<td>C</td>
<td>T</td>
<td>0.939</td>
<td>+</td>
<td></td>
<td>0.0007</td>
<td>0.0002</td>
</tr>
<tr>
<td>CHRN6</td>
<td>S43P</td>
<td>rs140930963</td>
<td>8</td>
<td>42620300</td>
<td>A</td>
<td>G</td>
<td>1.234</td>
<td>+</td>
<td></td>
<td>0.0034</td>
<td>0.0110</td>
</tr>
<tr>
<td>CHRN5</td>
<td>F76V</td>
<td>rs56351164</td>
<td>15</td>
<td>78873272</td>
<td>T</td>
<td>G</td>
<td>4.712</td>
<td></td>
<td></td>
<td>0.0021</td>
<td>0</td>
</tr>
<tr>
<td>CHRN5</td>
<td>V134I</td>
<td>rs2229961</td>
<td>15</td>
<td>78880752</td>
<td>A</td>
<td>G</td>
<td>6.721</td>
<td>+</td>
<td>++</td>
<td>0.0175</td>
<td>0.0042</td>
</tr>
<tr>
<td>CHRN5</td>
<td>K167R</td>
<td>rs80087508</td>
<td>15</td>
<td>78882233</td>
<td>A</td>
<td>G</td>
<td>5.403</td>
<td>+</td>
<td>++</td>
<td>0.0007</td>
<td>0.0158</td>
</tr>
<tr>
<td>CHRN5</td>
<td>L317V</td>
<td>rs11609178</td>
<td>15</td>
<td>78882682</td>
<td>G</td>
<td>C</td>
<td>2.262</td>
<td>+</td>
<td>++</td>
<td>0.0038</td>
<td>0.0014</td>
</tr>
<tr>
<td>CHRN5</td>
<td>L363O</td>
<td>rs79109919</td>
<td>15</td>
<td>78882821</td>
<td>T</td>
<td>A</td>
<td>4.767</td>
<td>+</td>
<td></td>
<td>0.0034</td>
<td>0.0433</td>
</tr>
<tr>
<td>CHRN5</td>
<td>D398N</td>
<td>rs16969968</td>
<td>15</td>
<td>78882925</td>
<td>A</td>
<td>G</td>
<td>3.766</td>
<td></td>
<td></td>
<td>0.2920</td>
<td>0.0520</td>
</tr>
<tr>
<td>CHRN5</td>
<td>E363D</td>
<td>rs61373491</td>
<td>15</td>
<td>78883895</td>
<td>G</td>
<td>C</td>
<td>3.285</td>
<td></td>
<td></td>
<td>0.0042</td>
<td>0</td>
</tr>
<tr>
<td>CHRN3</td>
<td>R37H</td>
<td>rs8192475</td>
<td>15</td>
<td>78911230</td>
<td>C</td>
<td>T</td>
<td>2.671</td>
<td>+</td>
<td>++</td>
<td>0.0457</td>
<td>0.0116</td>
</tr>
<tr>
<td>CHRN4</td>
<td>M457V</td>
<td>rs61373452</td>
<td>15</td>
<td>78917573</td>
<td>T</td>
<td>C</td>
<td>1.584</td>
<td></td>
<td></td>
<td>0.0192</td>
<td>0.0637</td>
</tr>
<tr>
<td>CHRN4</td>
<td>A435V</td>
<td>rs56317523</td>
<td>15</td>
<td>78921343</td>
<td>A</td>
<td>G</td>
<td>-0.549</td>
<td>+</td>
<td>++</td>
<td>0.0145</td>
<td>0.0007</td>
</tr>
<tr>
<td>CHRN4</td>
<td>A457T</td>
<td>-</td>
<td>15</td>
<td>78921344</td>
<td>T</td>
<td>C</td>
<td>3.053</td>
<td>+</td>
<td></td>
<td>0.0011</td>
<td>0</td>
</tr>
<tr>
<td>CHRN4</td>
<td>R421X</td>
<td>-</td>
<td>15</td>
<td>78921386</td>
<td>A</td>
<td>G</td>
<td>1.653</td>
<td>+</td>
<td>++</td>
<td>0.0014</td>
<td>0</td>
</tr>
<tr>
<td>CHRN4</td>
<td>T375I</td>
<td>rs61737499</td>
<td>15</td>
<td>78921523</td>
<td>G</td>
<td>A</td>
<td>2.262</td>
<td>+</td>
<td></td>
<td>0.0168</td>
<td>0</td>
</tr>
<tr>
<td>CHRN4</td>
<td>R349C</td>
<td>rs56235003</td>
<td>15</td>
<td>78921602</td>
<td>A</td>
<td>G</td>
<td>1.960</td>
<td>+</td>
<td>++</td>
<td>0.0842</td>
<td>0.0030</td>
</tr>
<tr>
<td>CHRN4</td>
<td>V311I</td>
<td>-</td>
<td>15</td>
<td>78921716</td>
<td>T</td>
<td>C</td>
<td>2.402</td>
<td>+</td>
<td></td>
<td>0.0062</td>
<td>0</td>
</tr>
<tr>
<td>CHRN4</td>
<td>S140G</td>
<td>rs56218866</td>
<td>15</td>
<td>78922229</td>
<td>C</td>
<td>T</td>
<td>2.269</td>
<td>+</td>
<td></td>
<td>0.0055</td>
<td>0.0410</td>
</tr>
<tr>
<td>CHRN4</td>
<td>R136Q</td>
<td>rs56095004</td>
<td>15</td>
<td>78922440</td>
<td>C</td>
<td>T</td>
<td>0.410</td>
<td></td>
<td></td>
<td>0.0124</td>
<td>0.0024</td>
</tr>
<tr>
<td>CHRN4</td>
<td>T91I</td>
<td>rs12914008</td>
<td>15</td>
<td>78923505</td>
<td>A</td>
<td>G</td>
<td>2.161</td>
<td></td>
<td></td>
<td>0.0391</td>
<td>0.0074</td>
</tr>
<tr>
<td>CHRN4</td>
<td>R18C</td>
<td>-</td>
<td>15</td>
<td>78933424</td>
<td>A</td>
<td>G</td>
<td>5.625</td>
<td></td>
<td></td>
<td>0.0014</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.1. Characteristics and Frequencies of Variants observed in the COGA dataset. For SIFT predictions, variants predicted to be damaging are depicted with +. For Polyphen, variants predicted to be “possibly damaging” or “probably damaging” are depicted with + and ++, respectively.
<table>
<thead>
<tr>
<th></th>
<th>CHRN3</th>
<th>CHRNA3</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Individuals</td>
<td>Non-Carriers</td>
<td>Carriers</td>
</tr>
<tr>
<td>Age (years)</td>
<td>37 ± 16</td>
<td>37 ± 15</td>
</tr>
<tr>
<td>Women/Men</td>
<td>2076/2048</td>
<td>25/19</td>
</tr>
<tr>
<td>Cigarettes per Day</td>
<td>21 ± 13</td>
<td>19 ± 12</td>
</tr>
<tr>
<td>Alcohol Dependent Cases / Controls</td>
<td>1796/2328</td>
<td>22/22</td>
</tr>
<tr>
<td>Alcohol Symptom Count</td>
<td>2.07 ± 2.4</td>
<td>2.27 ± 2.7</td>
</tr>
<tr>
<td>Cocaine Dependent Cases / Controls</td>
<td>677/3447</td>
<td>12/32</td>
</tr>
<tr>
<td>Cocaine Symptom Count</td>
<td>0.65 ± 1.8</td>
<td>1.3 ± 2.5</td>
</tr>
</tbody>
</table>

Table 3.2. Demographic and Phenotypic Characteristics of COGA European American Samples. Cigarettes per day (CPD), age, alcohol symptom count and cocaine symptom count values are mean ± s.d. P values for age and sex were calculated with a Fisher's exact test and Wilcoxon Rank Sum test, respectively. P values for cigarettes per day, alcohol dependence phenotypes, and cocaine dependence phenotypes were calculated using FamSKAT.
<table>
<thead>
<tr>
<th></th>
<th>CHRNA3</th>
<th>COGA African Americans</th>
<th>CHRNA3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Carriers</td>
<td>Carriers</td>
<td>P-value</td>
</tr>
<tr>
<td># of Individuals</td>
<td>1121</td>
<td>143</td>
<td>-</td>
</tr>
<tr>
<td>Age (years)</td>
<td>36 ± 13</td>
<td>37 ± 13</td>
<td>0.34</td>
</tr>
<tr>
<td>Women/Men</td>
<td>503/618</td>
<td>64/79</td>
<td>1</td>
</tr>
<tr>
<td>Cigarettes per Day</td>
<td>14 ± 11</td>
<td>14 ± 10</td>
<td>0.82</td>
</tr>
<tr>
<td>Alcohol Dependent Cases / Controls</td>
<td>530/590</td>
<td>74/69</td>
<td>0.32</td>
</tr>
<tr>
<td>Alcohol Symptom Count</td>
<td>2.99 ± 2.51</td>
<td>3.32 ± 2.56</td>
<td>0.13</td>
</tr>
<tr>
<td>Cocaine Dependent Cases / Controls</td>
<td>344/768</td>
<td>51/92</td>
<td>0.25</td>
</tr>
<tr>
<td>Cocaine Symptom Count</td>
<td>2.02 ± 2.91</td>
<td>2.30 ± 3.02</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Table 3.3. Demographic and Phenotypic Characteristics of COGA African American Samples. Cigarettes per day (CPD), age, alcohol symptom count and cocaine symptom count values are mean ± s.d. P values for age and sex were calculated with a Fisher’s exact test and Wilcoxon Rank Sum test, respectively. P values for cigarettes per day, alcohol dependence phenotypes, and cocaine dependence phenotypes were calculated using FamSKAT.
<table>
<thead>
<tr>
<th></th>
<th>CHRN3</th>
<th>CHRNA3</th>
<th></th>
<th>Non-Carriers</th>
<th>Carriers</th>
<th>P-value</th>
<th>Non-Carriers</th>
<th>Carriers</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Individuals</td>
<td>2036</td>
<td>11</td>
<td>-</td>
<td>1869</td>
<td>178</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>36 ± 5</td>
<td>36 ± 5</td>
<td>0.96</td>
<td>37 ± 5</td>
<td>36 ± 6</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women/Men</td>
<td>1247/789</td>
<td>41489</td>
<td>0.54</td>
<td>1142/727</td>
<td>113/65</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarettes per Day</td>
<td>18 ± 17</td>
<td>18 ± 19</td>
<td>0.99</td>
<td>19 ± 17</td>
<td>17 ± 17</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol Dependent Cases / Controls</td>
<td>464/1530</td>
<td>4/7</td>
<td>0.3</td>
<td>434/1395</td>
<td>34/142</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol Symptom Count</td>
<td>1.72 ± 1.66</td>
<td>1.81 ± 1.72</td>
<td>0.852</td>
<td>1.74 ± 1.67</td>
<td>1.50 ± 1.55</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cocaine Dependent Cases / Controls</td>
<td>131/1905</td>
<td>1/10</td>
<td>0.52</td>
<td>126/1743</td>
<td>6/172</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cocaine Symptom Count</td>
<td>0.43 ± 1.44</td>
<td>0.55 ± 1.81</td>
<td>0.8</td>
<td>0.45 ± 1.48</td>
<td>0.26 ± 1.12</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4. Demographic and Phenotypic Characteristics of COGEND European American Samples. Cigarettes per day (CPD), age, alcohol symptom count and cocaine symptom count values are mean ± s.d. P values for age and sex were calculated with a Fisher's exact test and Wilcoxon Rank Sum test, respectively. P values for CPD, alcohol dependence phenotypes, and cocaine dependence phenotypes were calculated using a Wilcoxon Rank Sum test.
<table>
<thead>
<tr>
<th></th>
<th>CHRN6B3</th>
<th>CHRNA3</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Carriers</td>
<td>Carriers</td>
<td>P-value</td>
<td>Non-Carriers</td>
<td>Carriers</td>
<td>P-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td># of Individuals</td>
<td>641</td>
<td>69</td>
<td>-</td>
<td>700</td>
<td>10</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>37 ± 6</td>
<td>33 ± 5</td>
<td>0.93</td>
<td>37 ± 5</td>
<td>38 ± 6</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women/Men</td>
<td>408/233</td>
<td>41/28</td>
<td>0.55</td>
<td>1142/727</td>
<td>113/65</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarettes per Day</td>
<td>16 ± 13</td>
<td>16 ± 12</td>
<td>0.63</td>
<td></td>
<td>16 ± 13</td>
<td>12 ± 10</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Alcohol Dependent Cases / Controls</td>
<td>66/575</td>
<td>13/56</td>
<td>0.06</td>
<td></td>
<td>78/622</td>
<td>1/9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Alcohol Symptom Count</td>
<td>1.35 ± 1.72</td>
<td>2.00 ± 2.34</td>
<td>0.004</td>
<td>1.43 ± 1.80</td>
<td>1.43 ± 2.6</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cocaine Dependent Cases / Controls</td>
<td>43/598</td>
<td>6/63</td>
<td>0.46</td>
<td></td>
<td>49/651</td>
<td>0/10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cocaine Symptom Count</td>
<td>0.71 ± 1.93</td>
<td>0.79 ± 2.10</td>
<td>0.77</td>
<td></td>
<td>0.42 ± 1.54</td>
<td>0 ± 0</td>
<td>0.39</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5. Demographic and Phenotypic Characteristics of COGEND African American Samples. Cigarettes per day (CPD), age, alcohol symptom count and cocaine symptom count values are mean ± s.d. P values for age and sex were calculated with a Fisher's exact test and Wilcoxon Rank Sum test, respectively. P values for CPD, alcohol dependence phenotypes, and cocaine dependence phenotypes were calculated using a Wilcoxon Rank Sum test.
Figure 3.4. Distribution of DSM-IV Alcohol or Cocaine Dependence Symptom Count among carriers or non-carriers of A) any CHRNA3 variant or B) CHRNA3 H329Y among European American individuals in the COGA dataset
Figure 3.5. Distribution of CPD by number of DSM-IV cocaine dependence symptoms endorsed. Each red circle is proportional to the number of individuals with a given value of CPD calculated for each number of endorsed DSM-IV cocaine dependence symptoms endorsed. Proportions per number of symptoms endorsed was used to allow easier comparison of proportions across the range of symptom count despite differing numbers of individuals.
Figure 3.6. Distribution of DSM-IV cocaine symptom count by number of DSM-IV alcohol dependence symptoms endorsed. Each red circle is proportional to the number of individuals with a given number of cocaine symptoms endorsed calculated for each number of endorsed DSM-IV alcohol dependence symptoms endorsed. Proportions per number of symptoms endorsed was used to allow easier comparison of proportions across the range of symptom count despite differing numbers of individuals.
Figure 3.7. Distribution of CPD by number of DSM-IV alcohol dependence symptoms endorsed. Each red circle is proportional to the number of individuals with a given value of CPD calculated for each number of endorsed DSM-IV alcohol dependence symptoms endorsed. Proportions per number of symptoms endorsed was used to allow easier comparison of proportions across the range of symptom count despite differing numbers of individuals.
Chapter 4. Functional characterization of rare variants in *CHRNB4* to improve the association between *CHRNB4* and nicotine dependence

**ABSTRACT**

Smoking is the leading cause of preventable death worldwide. As such, effort has been devoted to determining what genetic variants contribute to smoking risk. Genome-wide association studies have identified variants in nicotinic acetylcholine receptor genes that contribute to nicotine dependence risk and follow up studies of these variants suggest that the same variants may play a role in multiple substance use disorders including alcohol and cocaine dependence. We previously undertook pooled sequencing of the coding regions and flanking sequence of the *CHRNA5, CHRNA3, CHRNB4, CHRNA6* and *CHRN B3* genes and found that rare missense variants in *CHRNB4* are associated with reduced risk of nicotine dependence. Recent *in vivo* studies have also implicated *CHRNB4* in nicotine addiction: Mice over-expressing Chrnb4 have been shown to display a strong aversion to nicotine that can be reversed by virally mediated expression of Chrna5 D398N variant, a variant that reduces nicotinic receptor function, in the medial habenula. It has yet to be determined if rescue with wild-type Chrna5 is possible, however. Together, these results highlight the importance of understanding the function of variants in nicotinic acetylcholine receptors and *CHRNB4* in particular. Our sequencing efforts, led to the identification of 18 rare (<5%) non-synonymous variants in *CHRNB4* in a total of 2067 individuals from the Collaborative Study of the Genetics of Nicotine Dependence (COGEND) cohort. In order to investigate the effect of these variants, the normal or variant β4 subunit was co-expressed with normal α3 subunit in
human embryonic kidney 293 cells. Whole-cell voltage-clamp was used to obtain acetylcholine and nicotine concentration–response curves for each subunit combination. qRT-PCR, western blots and cell-surface ELISAs were performed to assess differences in mRNA production, overall CHRN4 protein expression and cell-surface specific protein expression levels, respectively. Analysis of these data revealed several variants that affected nicotine EC$_{50}$, acetylcholine EC$_{50}$, maximal response to nicotine, maximal response to acetylcholine or nicotine efficacy. Nearly all variants tested lowered cell-surface expression. These results highlight the need for detailed functional analysis of variants using multiple modalities and the advantages to integrating such data into human genetics.

**INTRODUCTION**

Nicotinic acetylcholine receptors (nAChRs) are pentameric ligand-gated ion channels formed from numerous combinations of receptor subunits, each encoded by a separate gene in the human genome. Neuronal nAChR α subunits are encoded by six genes (CHRNA2-A7, CHRNA9 and CHRNA10) and the β subunits by three genes (CHRNB2-B4). A minimum of two α and two β subunits are required to form functional heteromeric receptors. This is because acetylcholine (ACh) binding occurs at the interface between α and β subunits. The only subunit capable of forming homomeric receptors is CHRNA7. Certain combinations of receptor subunits are more common in the human central nervous system (CNS). There exists profound regional specificity with regard to subtype expression in the mammalian CNS. The expression of the α3β4* subtype (the asterisk denotes any one of multiple accessory subunits), for instance, is limited for the most part
to autonomic and sensory ganglia, medial habenula, fasciculus retroflexus and the interpeduncular nucleus (IPN), while α4β2* receptors can be found almost ubiquitously throughout the brain.141-143

Recently, work has been done that suggests that the α3β4* subtype of nicotinic receptors may largely contribute to nicotine related behaviors by activating the habenulo-interpeduncular pathway. The activity of the habenulo-interpeduncular pathway was recently shown to be increased during nicotine withdrawal and is involved in nicotine self-administration. Inhibiting α3β4* nicotinic receptors in the habenula reduces nicotine self-administration while blocking α3β4* nicotinic receptors in the IPN increases nicotine self-administration.144 Knocking out CHRNA5 results leads to dramatic increases in nicotine intake, even at high concentrations, and is reversible upon lentiviral expression of CHRNA5 in the medial habenula.77 Further, mice over-expressing Chrnb4 were shown to display a strong aversion to nicotine, which was reversed upon virus-mediated expression of the a5 D398N variant in the medial habenula.39 This work suggests that aversion of nicotine in mouse is regulated by the balanced activity of β4 and α5 nicotinic receptor subunits in the medial habenula. Additionally, a recent association study of smoking cessation has shown that variants in CHRNA4 may decrease craving and withdrawal symptoms.91 These findings suggest that multiple nicotinic receptor subunit genes likely play a role in the development and maintenance of nicotine dependence and that variants in these genes may have different and in some cases opposing functional effects in humans and mice.

Several variants in the CHRNA4 gene have been shown to affect receptor function using one of these in vitro assays.88 The CHRNA4 variants T91I, R136W, S140G, and
M467V were each expressed heterologously in Xenopus oocytes and studied using the two-electrode voltage clamp. These experiments revealed that the R136W and M467V forms of the receptor had higher sensitivity to acetylcholine and lower EC_{50} than the wild-type. The T91I variant was shown to have lower sensitivity to acetylcholine and a larger EC_{50}. Each of the T91I, R136W and M467V variants also showed greater desensitization after long-term exposure to low-dose acetylcholine. These results suggest that variants in nicotinic receptor genes alter receptor function in several ways, but also that more work needs to be done to fully understand the contribution of variation in the coding regions of CHRNA4 and other nicotinic receptor genes.

Recently, multiple variants in or near nicotinic acetylcholine receptor genes have been found to be associated with nicotine related behavior in humans. A non-synonymous change (rs16969968/D398N) in CHRNA5 is the most strongly associated SNP in several genome-wide association (GWA) studies of nicotine dependence. This variant was also shown to affect risk for cocaine dependence, although the risk allele for nicotine dependence decreases risk for cocaine dependence^{61,62}. Additionally, variants near CHRNA5 previously shown to alter CHRNA5 mRNA expression in vivo, alter risk for both nicotine and alcohol dependence^{41,45}. A group of common SNPs near the CHRN3-CHRNA5 gene cluster were also shown to affect cigarette consumption in a recent GWAS^{43}. Sequencing of the neuronal nicotinic receptor genes in large cohorts of nicotine dependent cases and controls has also found associations between variants in two nicotinic receptor genes, CHRNA4 and CHRB4, though we do not replicate the previous findings in CHRNA4^{52,138}. In order to investigate the functional consequences of the variants associated with nicotine dependence from our study^{138}, we constructed vectors
harboring variant CHRN\textsubscript{B4} alleles, expressed the proteins in Hek293 cells and examined receptor function using several assays. These results were then incorporated into genetic association analyses. The goal of this chapter is to characterize the functional impact of rare variants in the \textit{CHRN\textsubscript{B4}} and \textit{CHRNA3} genes observed from sequencing a cohort of nicotine dependent individuals and non-dependent smoking controls in order to investigate the hypothesis that modeling the specific functional effects of variants in these genes, and genes more generally, could be used to improve association analyses of such variants with disease phenotypes.

\section*{METHODS}

\textit{Generation and expression of constructs}

Full-length coding sequences for the human nicotinic $\alpha$3 (NP_000734.2) and $\beta$4 (NP_000741.1) subunits were kindly provided by Dr. J. Lindstrom (University of Pennsylvania, Philadelphia, PA). Subunits were sub-cloned into the pcDNA3 vector (Invitrogen, San Diego, CA). Mutations were introduced using the QuikChange kit (Stratagene, San Diego, CA). The FLAG epitope \([\text{DYKDDDDK}^{145}]\) was introduced into $\alpha$4 between the 6 and 7 positions of the mature polypeptide using QuikChange. For each construct the entire subunit coding region was sequenced to verify that only the desired mutation had been introduced.

\textit{Cell culture and transfection.}\n
HEK 293 cells (American Tissue Culture Collection, Gaithersburg, MD) were maintained in a mixture of Dulbecco’s modified Eagle’s medium and Ham’s F12 (1:1,
also containing L-glutamine and 15 mM HEPES), with 10% fetal bovine serum (Hyclone, Logan, UT), penicillin (100 units/ml) and streptomycin (100 ug/ml) in a humidified atmosphere containing 5% CO2 at 37°C. Cells were re-plated in the same medium the day before transfection.

For physiological and surface ELISA studies, subunits were transfected at a 1:1 mass ratio using Effectene (QIAGEN, Valencia, CA) according to the manufacturer’s instructions. Briefly, 3 µg of cDNA per well of a 24-well dish was mixed with the Enhancer and the Effectene Transfection Reagent. The cells were incubated with the mix for 6 to 18 h. Electrophysiological experiments commenced the following day, while ELISAs were performed on the third day after transfections.

**Cell-Surface ELISAs**

Surface ELISA assays were performed basically as described \(^{146,147}\). Cells were plated in 24 well tissue culture plates at about 100,000 cells/well. The next day cells were transfected as described. In each experiment, a negative control (empty pcDNA3) and a positive control (wild-type subunits) were performed. Five wells were transfected with each subunit combination in each experiment; 3 were used for ELISA and 2 for a protein assay. For the ELISA assay, cells were rinsed in phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na\(_2\)HPO\(_4\), and 1.4 mM KH\(_2\)PO\(_4\), pH 7.3) then blocked for 30 min at room temperature with 4% powdered milk in PBS (milk-PBS). To detect α3 mAb 35 was used as primary antibody (Sigma-Aldrich, St. Louis MO; diluted 1:400 in milk-PBS) for 1 h at room temperature. Cells were then washed twice with milk-PBS and incubated with anti-rat IgG peroxidase-conjugated goat antibodies (Sigma, A9037; 1:100
dilution in milk/PBS) for 1 h and washed with PBS four times. To detect FLAG-tagged α4, cells were incubated with M2 antibody (Sigma-Aldrich, 2 µg/ml in milk-PBS) as described and then incubated with anti-mouse IgG peroxidase-conjugated goat antibodies (Sigma, A9037; 1:100 dilution in milk/PBS) for 1 h. After incubation with secondary antibody, cells were washed 4 times with PBS, and then assayed using the 1-Step Ultra TMB-ELISA kit (Pierce, Rockford, IL). Absorbance was read at 405 nm using a microplate reader (iMark, Biorad, Hercules CA). Total cell protein was assayed from wells that had been maintained in milk-PBS, then washed twice with PBS before assay by a bicinchoninic acid method (Pierce). For each experiment, the ELISA signal was obtained from triplicate wells and the cell protein from duplicate wells.

The surface ELISA data were analyzed as follows. The machine background was subtracted from each OD reading, then the OD readings were divided by the protein for that subunit combination. The normalized value for the negative control (pcDNA3) for that experiment was then subtracted from all values. Finally, to control for variation in expression between experiments, the subtracted expression levels were normalized to the positive control (wild-type) value for that experiment. The final value gives an estimate of the relative expression where a value of 1 indicates identical expression to wild-type subunits.

**Western Blots**

To measure *CHRNB4* protein levels in HEK293 cell lines transiently transfected with pcDNA3-empty, pcDNA3-*CHRNA3*-Wild-Type and pcDNA3-*CHRNB4*-Wild-Type, or pcDNA3-*CHRNA3*-WT and pcDNA3-*CHRNB4*-Mutant plasmids, we performed
western blots on cell lysates using a poly-clonal \textit{CHRNB4} antibody generously provided by Dr. Cecelia Gotti. Cells were seeded into 6-well lysine-coated plates and cultured in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine, and penicillin/streptomycin. Cells were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer’s protocol incubated in a humidified atmosphere containing 5% CO$_2$ at 37°C. After two days of growth, each well of the 6-well plate was used to create a cell-lysate. Cells were lysed with 150ul of lysis buffer (50 mM Tris, 150 mM NaCl, 2 mM EDTA, 0.25% NP40, 1% TritonX). Total protein concentration was then measured by BCA assay for each cell lysate. 20ug of protein from each lysate was then incubated at 95°C for 5 minutes in 1x lamellae buffer (0.25 M Tris, 8% SDS, 40% glycerol, 0.01% bromophenol blue dye and 20% β-mercaptoethanol). Denatured samples were then loaded onto a 4-20% Criterion gel in TG-SDS buffer (0.01% SDS, 25 mM Glycine, 2.5 mM Tris) and run at 125V for 90 minutes. Protein in the gel was then transferred to a nitrocellulose membrane in TG-SDS buffer containing 20% methanol overnight at 4°C. Blots were then incubated in TG-SDS containing 4% milk for 25 minutes at room temperature, incubated at room temperature with a primary \textit{CHRNB4} poly-clonal antibody for 90 minutes, rinsed 3x with PBS containing 1% Triton-X for 5 minutes, incubated with a horseradish peroxidase conjugated secondary antibody, washed 3x with PBS containing 1% Triton-X for 5 minutes and finally incubated with the horseradish peroxidase substrate 3, 3, 5, 5, 0-tetramethylbenzidine before images were taken allowing for 5 minutes of exposure using a Syngene western blot imager.
mRNA Expression

To measure CHRN4 mRNA production in HEK 293 cells transiently transfected with pcDNA3-empty, pcDNA3-CHRHA3-Wild-Type and pcDNA3-CHRN4-Wild-Type, or pcDNA3-CHRHA3-WT and pcDNA3-CHRN4-Mutant plasmids, RNA was extracted from cell lysates with an RNeasy kit (Qiagen). Extracted RNA (10 ug) was then converted to cDNA using the High Capacity cDNA Reverse Transcriptase kit (ABI). CHRN4 mRNA expression was then measured using sybrgreen (Invitrogen) using one primer spanning exons 3 and 4 and another primer spanning exons 4 and 5 to ensure only cDNA was amplified.

Whole-cell patch clamp

Cells were plated in 35 mm tissue culture dishes, and maintained and transfected as described. For recordings, cells were rinsed with recording bath solution (140 mM NaCl, 5 mM KCl, 1 mM MgCl₂, 2 mM CaCl₂, 10 mM glucose, and 10 mM HEPES, pH 7.4) and cells expressing high levels of surface receptors were identified by a bead-binding technique⁸⁶. We used mAb35 (Sigma-Aldrich) to identify α3 and mAb199 for α4 (Sigma-Aldrich). Antibody was adsorbed to immunobeads with a covalently attached secondary antibody (Invitrogen). The cells were incubated with a suspension of beads for 5 to 10 min with gentle shaking, and cells expressing surface receptors were identified from the presence of beads bound to the cell. This greatly enhanced our ability to identify cells expressing measurable numbers of receptors.

Macroscopic currents were recorded using whole-cell voltage clamps described before⁸⁶. A diagram of the whole-cell voltage clamp technique can be seen in figure
4.1. The pipette (intracellular) solution contained 140 mM CsCl, 4 mM NaCl, 4 mM MgCl₂, 0.5 mM CaCl₂, 5 mM EGTA, and 10 mM HEPES, pH 7.4. The drugs were applied through the bath using an SF-77B fast perfusion stepper system (Warner Instruments, Hamden, CT). Currents were recorded using an Axopatch 200 amplifier (Molecular Devices, Union City, CA). The cells were clamped at -60 mV. All experiments were carried out at room temperature (19–22°C). The current traces were low-pass filtered at 2 kHz and digitized at 10 kHz. The analysis of whole-cell currents was carried out using the pClamp 9.0 software package (Molecular Devices).

The basic parameter measured was the peak response to a given concentration of acetylcholine (ACh) or nicotine. Four second pulses of agonist were applied at intervals of 30 seconds with continuous application of bath solution between pulses to allow recovery from desensitization. Concentration-response relationships were constructed by applying pulses of differing agonist concentration. The data from a cell were analyzed by fitting with the Hill equation: \( Y(X) = a/(1+(X/k)^n) \) where \( Y(X) \) is the response to a concentration \( X \), and the parameters \( a \) (maximal response), \( k \) (EC\(_{50}\) or concentration giving half-maximal response) and \( n \) (Hill coefficient) were determined by non-linear regression using SigmaPlot (Systat Software, Chicago IL). To combine data, the relationships were scaled by the fit maximum. In some cases, it was clear that high concentrations of agonist produced a reduced response, likely due to open-channel block by the agonist\(^ {148-150} \). Accordingly, responses to high concentrations that produced a lower response than responses to lower concentrations were not included in the fit.

Concentration-response data for ACh and nicotine were obtained on different cells, to minimize the duration of whole-cell recording. Previous work has demonstrated
that there can be changes in the peak response and/or the desensitization properties of neuronal nicotinic receptors over the time of whole-cell recording\(^{150,151}\). Accordingly, each cell was tested with a high concentration of ACh (usually 1 mM), and this value was used to normalize the nicotine concentration-response relationship to the overall averaged ACh concentration-response relationship for that particular subunit combination.

Similarly, some cells were tested with single applications of a high concentration of ACh and a high concentration of nicotine to obtain estimates for relative maximal responses. In this case, the relative maximal response to nicotine was normalized to both the mean concentration-response data for nicotine and for ACh for that subunit combination.

The fact that cells were not randomly selected means that the average response to a maximal concentration of ACh does not give a direct estimate of the number of activatable receptors on an average cell and so is not comparable to the surface ELISA signal. There was a general correlation between the average maximal response and the ELISA signal.

**Drugs, data presentation and statistics.**

Unless otherwise noted all chemicals used were obtained from Sigma-Aldrich. Data are presented as mean ± SE (number of observations). For ELISA studies results from each experiment constituted one observation; data were obtained in triplicate in each experiment. For physiological results each cell constituted an observation (i.e. EC\(_{50}\), or maximal response).

Parameters from fitting the Hill equation were excluded from analysis if the standard error of any fit parameter for that cell (as estimated by the fitting program) was
60% or more than the best fitting value. Data from ELISA experiments were excluded if the expression for the positive control was less than 3-times pcDNA3, indicating a failed transfection.

Basic statistical computations were made with Excel (Microsoft, Redmond WA). ANOVA was performed using Stata (StataCorp, College Station TX). Figures were prepared with SigmaPlot.

The homology model was made by threading the $\alpha_3$ and $\beta_4$ subunits onto the $\alpha$ and $\delta$ subunits respectively of the Torpedo marmorata cryoelectron microscopic structure (PDB 2BG9) using the SWISS-MODEL web tool (http://swissmodel.expasy.org/). Structures were visualized and displays generated using Chimera 1.6.2 (http://www.cgl.ucsf.edu/chimera).

**Samples and Phenotype**

DNA samples were collected as part of the Collaborative Genetic Study of Nicotine Dependence (COGEND). All members of the COGEND sample underwent a semi-structured interview, which assessed smoking behavior, other substance use and comorbid psychiatric conditions. The COGEND sample includes 710 African Americans (461 nicotine dependent (ND) cases and 249 smokers with no symptoms of dependence (controls)) and 2055 European Americans (1062 ND cases and 993 controls). As these individuals were ascertained to study nicotine dependence, Fagerström Test of Nicotine Dependence (FTND) scores (Range= 1-10) were required to be $\leq 1$ for controls and $\geq 4$ for cases. In all cases lifetime maximum FTND score was used. All members were required to have smoked at least 100 cigarettes in their lifetime. Because in our original analyses
the association between rare missense variants and CPD in African Americans was the strongest, we used this as our primary sample and phenotype when assessing whether functional characterization of alleles could improve observed associations. As linear regression was used and requires normally distributed variables, we log-transformed CPD such that it was normally distributed. A total of 352 African Americans (176 ND cases and 176 smoking controls) and 400 European Americans (200 ND cases and 200 smoking controls) were sequenced in the original study. Follow up genotyping of SNPs identified and validated in the sequenced individuals was done in the remaining portion of the COGEND sample (310 African Americans and 1662 European Americans).

**Association Analysis**

Association analyses were performed in R using linear regression incorporating age and sex as covariates. To analyze carrier status, individuals were coded as either 0 or 1 depending on whether they carried 0 or \( \geq 1 \) missense variants at any position in \( CHRN B4 \). Carrier status was tested in a linear regression against logCPD with age and sex as covariates. For all variant combinations of multiple \( CHRN B4 \) variants occurring in at least one individual, plasmids were created that harbored both variants. To analyze the function-weighted carrier status, individuals were coded as either 0 or 1 depending on whether they carried 0 or \( \geq 1 \) missense variants at any position in \( CHRN B4 \) and then the person’s code was multiplied by the normalized value for each of the tested parameters (Acetylcholine \( EC_{50} \), Nicotine \( EC_{50} \), acetylcholine maximal response, nicotine maximal response and nicotine efficacy). These values were then used as the predictor in a linear regression using age and sex as covariates.
RESULTS

Pharmacological Profiles of Variant containing α3β4 Receptors

A total of 10 rare missense variants were identified in *CHRNB4* as part of the previous sequencing projects. These variants can be found in many domains of the mature protein, but many occur within the first extracellular domain (Figure 4.40). To test the electrophysiological properties of β4 containing nicotinic receptors constructed to harbor these variant alleles, wild-type α3 and variant containing β4 receptor subunits were co-expressed in HEK 293 cells. Responses from cells were measured upon exposure to between 1 and 10,000 µM ACh and subsequently to 1 to 300 µM nicotine. Representative traces and resultant concentration response curves can be seen in figure 4.41. Concentration response curves for nicotine for each variant can be seen in figures 4.9-4.23. Concentration response curves for acetylcholine can be seen in figures 4.24-4.39. Several variants show altered response to nicotine. Specifically, nicotine EC₅₀ was altered for two variants (M456V and K57E) and nicotine efficacy was altered for one *CHRNB4* variants (T91I) when combined with the α3 variant R37H, a variant that almost always accompanies T91I in individuals of both African and European descent. Values and statistical significance for all electrophysiological assays can be seen in Figure 4.41, 4.42, 4.45 and tables 4.1, 4.2 and 4.3. No variants observed during sequencing significantly affected any electrophysiological parameters when expressed with α4 subunits rather than α3.
**Cell-Surface ELISAs**

In order to determine the extent to which mutations in *CHRNB4* and *CHRNA3* alter the steady-state level of nAChR protein expressed on the cell-surface, we performed ELISAs on cultured HEK 293 cells two days after transient transfection with constructs expressing wild-type *CHRNA3* and either wild-type *CHRNB4* or mutant *CHRNB4*. A diagram of the technique can been seen in figure 4.2. Cell-surface expression was decreased for nearly all variants tested (Figure 4.44, Table 4.3 and Table 4.4.). The S140G variant is predicted to abolish glycosylation at amino-acid position 138. The S is third amino acid in the consensus glycosylation signal NXS, where X can be any amino acid. We hypothesized that the absence of glycosylation at this site results in low cell-surface protein expression of this variant. First, to test this hypothesis, we created a *CHRNB4* expressing plasmid containing an alanine (A) at position 138, rather than the asparagine (N) required for glycosylation at this site as well as a plasmid containing a threonine (T) at position 140, predicted to retain glycosylation but change the amino acid at position 140. Cell-surface ELISAs demonstrated cell-surface protein levels of the N138A variant to be low and similar in level to the S140G variant while the S140T variant showed cell-surface protein levels more closely resembling wild-type *CHRNB4*, albeit reduced (Figure 4.44, Table 4.3 and Table 4.4.). These data suggest that at least for the S140G variant, it is ablation of glycosylation at this position that explains the reduction in cell-surface protein expression of S140G variant receptors. We also observe a significant correlation between relative cell-surface ELISA results when variant containing subunits are expressed with α3 or α4 subunits (Figure 4.46) and a significant
correlation between relative maximal response to acetylcholine and relative cell-surface ELISA levels when expressed with α3 but not α4 (Figure 4.47).

**Western Blots**

In order to determine if overall *CHRNB4* protein, both intra-cellular and cell-surface, was altered by the introduction of mutation, we performed western blots on total cell lysates from HEK 293 cells transiently transfected with wild-type *CHRNA3* containing plasmids and either wild-type or mutant *CHRNB4* containing plasmids. Analysis of the band intensity for each mutant protein provided no evidence of altered total *CHRNB4* expression for any of the variant plasmids (Figure 4.3, 4.4, 4.5 and Figure 4.8). Of note, each of the variants predicted to alter glycosylation at amino-acid position 138 (S140G, N138A and S138A/S140G) show shifts in their full-length protein band in the expected direction assuming a disruption of glycosylation (Figure 4.3). Further, these differences in apparent molecular mass were normalized upon incubation with PNGase to remove all glycosylation (Figure 4.6 and Figure 4.7).

**Association Analyses**

To determine if incorporation of results from our functional analyses of all variants known to exist in our study population could be used to improve power to detect an association between nicotine related behaviors and variants in *CHRNB4*, we created a genotype model weighted based on the predicted maximal nicotine response of each of the variants. We used each parameter estimated from a variant’s concentration response curves for nicotine and acetylcholine as well as cell-surface protein level to create quantitative measures of receptor function that we could assign to each individual based
on the alleles they harbor. We first performed an analysis simply using carrier status at any non-synonymous site in *CHRNB4* as the predictor variable in a linear regression with log transformed lifetime maximum number of cigarettes smoked per day (logCPD) as the response variable and including age and sex as covariates. Using all missense variants, there is no significant association between carrying at least one missense variant in *CHRNB4* and logCPD ($\beta = -0.038$, $p = 0.585$, $r^2 = 0.0009$) (Figure 4.48). We then took the subset of missense variants that occurred at genomic positions with at least suggestive levels of cross-species conservation (vertebrate PhyloP scores > 2). This reduced the number of variants from ten to six. We observed a modest, but significant association between carrying at least one missense variant at a conserved site in *CHRNB4* and number of cigarettes smoked/day ($\beta = -0.24$, $p$-value $= 0.008$, $r^2 = 0.009$) (Figure 4.49). We then weighted each variant genotype by the value for each of the estimated electrophysiological parameters (nicotine EC$_{50}$, acetylcholine EC$_{50}$, maximal response to nicotine, maximal response to acetylcholine, nicotine efficacy) or relative cell-surface expression level. The results of this weighting procedure can be seen in Figures 4.50-4.55. Weighting in this way by acetylcholine EC$_{50}$ explained substantially more phenotypic variance and produced an association that would be considered genome-wide significant for gene-based associations assuming a multiple test correction for 20,000 genes ($\beta = -0.006$, p-value $= 6 \times 10^{-6}$, $r^2 = 0.029$). These results suggest that substantial power can be gained by incorporating knowledge of the functional effects of variants into genetic association analyses.

**DISCUSSION**
The goal of this chapter was to characterize the functional impact of rare variants in the \textit{CHRNB4} and \textit{CHRNA3} genes observed from sequencing a cohort of nicotine dependent individuals and non-dependent smoking controls in order to investigate the hypothesis that modeling the specific functional effects of variants in these genes, and genes more generally, could be used to improve association analyses of such variants with disease phenotypes. To do so we performed electrophysiological experiments, cell-surface ELISAs, western blot assays and qRT-PCR to thoroughly describe the effects of variants in these genes on receptor expression and function. Some variants had little effect on receptor function. However, most variants had some measurable effect on cell-surface protein expression compared to wild-type. As several variants have strong effects on cell-surface protein expression despite having no measurable effect on smoking behavior, it seems reasonable that alterations to protein function are better indicators of whether or not a variant will affect behavior. Additionally, as the intracellular protein level is generally much higher than at the cell-surface, individuals heterozygous for mutations that cause lower cell-surface expression when expressed alone may be compensated for by their non-variant allele. While many variants perturbed the cell-surface protein levels of the receptors they were incorporated into, none substantially altered total \textit{CHRNB4} protein expression. Additionally, only a small number of the characterized variants significantly affected aspects of nicotine or acetylcholine response. The significant relationship between small perturbations of protein function when taken together and a complex human behavior like nicotine dependence suggests that precise measurements of allele function are critical. These findings also suggest that it is important to investigate multiple avenues of protein function alteration to fully
characterize the impact of variants. Though mildly correlated, the electrophysiological parameters used in this study likely have distinct effects on cognition and the role of exogenous chemicals like nicotine in brain chemistry.

Overall, the results from our electrophysiological assays are consistent with findings reported previously by others. These experiments revealed that the R136W and M467V variants had higher sensitivity to acetylcholine and lower EC$_{50}$ than the wild-type and that the T91I variant had lower sensitivity to acetylcholine and a larger EC$_{50}$. We confirm these results, despite having performed the experiments in transiently transfected HEK 293 cells and $CHRNB4$ variant subunits were expressed with $CHRNA3$ rather than $CHRNA4$. Though not significant, the Liang et al. study also found the S140G variant to increase acetylcholine EC$_{50}$, consistent again with our findings.

We hypothesized that functional characterization of variants could improve power to detect associations between assayed rare variants in a gene-based burden-like test. In order to test this, we performed a regression based rare variant burden test in which variants were weighted in proportion to their effect on each of the following concentration response curve parameters: acetylcholine EC$_{50}$, nicotine EC$_{50}$, acetylcholine maximal response, nicotine maximal response, nicotine efficacy and cell-surface expression. We find that acetylcholine EC$_{50}$ was the best predictor of behavior as measured by log(CPD). No other parameter improved our ability to observe an association at this locus. Overall, our results suggest that this approach of incorporating functional information can improve power to detect associations if relevant parameters are measured and that methods of assaying the functional impact of variants across the genome will likely greatly improve gene discovery in the era of whole-exome and whole-
genome sequencing. The reduction in power resulting from large proportions of variants with little or no impact on protein function or mixtures of protective and risk variants being included in gene-based burden tests is substantial and will have to be addressed if we hope to understand the full scope of variation impacting common complex diseases and traits.

In conclusion, we find that the majority of missense variants in *CHRNB4* and *CHRNA3* alter receptor function or receptor cell-surface expression. Further, these findings may be used to inform genetic association analyses to improve power to detect signals often obscured by more common neutral or near neutral alleles. These findings highlight the need for the integration of functional studies into the findings of human genetics.
Figure 4.1. Mechanics of whole-cell voltage clamp. First a glass pipette with slight negative pressure is applied to the outside of a cell. A brief burst of suction is the applied to break the cell membrane, allowing the inside of the pipette to be continuous with the inside of the cell. The degree to which the cell is depolarized upon treatment with an agonist or antagonist is then monitored. Figure is taken from: http://www.leica-microsystems.com/science-lab/the-patch-clamp-technique/.
Figure 4.2. Schematic of Cell-Surface ELISA assay Mechanics. Cells transiently transfected with plasmids expressing wild-type CHRNA3 in addition to either wild-type CHRNβ4 or variant containing CHRNβ4 are incubated with primary antibody against the extracellular domain of CHRNA3 derived from rat (purple). The cultures are then incubated with a horseradish peroxidase conjugated anti-rat secondary antibody (blue) and then allowed to induce color change in tetramethylbenzidine (white circles and orange circles). Figure adapted from.
Figure 4.3. Western Blot of S140G and related variants. N138A, S140G and the doubly mutated plasmid containing both N138A and S140G, all predicted to affect glycosylation at residue 138, run more quickly through a polyacrylamide gel than do either S140T or wild-type proteins, both predicted to permit glycosylation at residue 138.
Figure 4.4. Total $CHRNB4$ protein level for variant residues 57-136. Images are representative of 6 replicate transfections averaged for statistical analyses. Protein levels were normalized to beta-tubulin levels (bottom).
Figure 4.5. Total CHRN4 protein level for variant residues 296-467. Images are representative of 6 replicate transfections averaged for statistical analyses. Protein levels were normalized to beta-tubulin levels (bottom).
Figure 4.6. Total and PNGase digested CHRN4 protein levels and size shifts for variant residues related to glycosylation. Incubation with PNGase F normalizes the distance run on a polyacrylamide gel for variants predicted to affect glycosylation and those that are not.
Figure 4.7. Total and PNGase digested *CHRNB4* protein levels and size shifts for variant residues related to glycosylation. Incubation with PNGase F normalizes the distance run on a polyacrylamide gel for variants predicted to affect glycosylation and those that are not.
Figure 4.8. Western blot quantification. Replicates across multiple days of transfection were averaged and normalized to wild-type *CHRN*4 protein level.
Figure 4.9. Nicotine concentration response curve for wild-type and R37H containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.10. Nicotine concentration response curve for wild-type and K57E containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.11. Nicotine concentration response curve for wild-type and T91I containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.12. Nicotine concentration response curve for wild-type and R136W containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.1. Nicotine concentration response curve for wild-type and R136Q containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.14. Nicotine concentration response curve for wild-type and S140G containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.15. Nicotine concentration response curve for wild-type and R349C containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.16. Nicotine concentration response curve for wild-type and T375I containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.17. Nicotine concentration response curve for wild-type and M456V containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.18. Nicotine concentration response curve for wild-type and M467V containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.19 Nicotine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of R37H \textit{CHRNA3} and T91I \textit{CHRNB4}. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.20. Nicotine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of wild-type CHRNA3 and a 1:1 mixture of M467V CHRNB4 and R136W CHRNB4. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.21. Nicotine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of wild-type CHRNA3 and a 1:1 mixture of M467V CHRNB4 and S140G CHRNB4. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.22. Nicotine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of wild-type *CHRNA3* and *CHRNB4* with both the M467V and R136W variants. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.23. Nicotine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of wild-type *CHRNA3* and CHRNA4 with both the M467V and S140G variants. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.24. Acetylcholine concentration response curve for wild-type and *CHRNA3* R37H containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.25. Acetylcholine concentration response curve for wild-type and \textit{CHRN}B4 K57E containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.26. Acetylcholine concentration response curve for wild-type and *CHRNB4* K57E containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.27. Acetylcholine concentration response curve for wild-type and *CHRN4* R136W containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.28. Acetylcholine concentration response curve for wild-type and CHRN4 R136Q containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.2. Acetylcholine concentration response curve for wild-type and \( CHRN B4 \) S140G containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.30. Acetylcholine concentration response curve for wild-type and *CHRNB4* G296S containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.31. Acetylcholine concentration response curve for wild-type and $CHRNB4$ R349C containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.3. Acetylcholine concentration response curve for wild-type and *CHRNB4* T375I containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.33. Acetylcholine concentration response curve for wild-type and \textit{CHRNB4} M456V containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.3. Acetylcholine concentration response curve for wild-type and \textit{CHRNB4 M467V} containing nicotinic receptors. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.35. Acetylcholine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of R37H CHRNA3 and T91I CHRN4B4. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.36. Acetylcholine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of wild-type CHRNA3 and a 1:1 mixture of M467V CHRNB4 and R136W CHRNB4. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.37. Acetylcholine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of wild-type CHRNA3 and a 1:1 mixture of M467V CHRN4 and S140G CHRN4. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.38. Acetylcholine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of wild-type CHRNA3 and CHRNA4 with both the M467V and R136W variants. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4.39. Acetylcholine concentration response curve for wild-type nicotinic receptors and nicotinic receptors composed of wild-type CHRNA3 and CHRNA4 with both the M467V and S140G variants. Variant (red curve) response parameters were then estimated and compared to wild-type (black curve) parameters.
Figure 4. Location of variants in a homology model of the α3β4 receptor. A homology model of the receptor was made by threading the α3 and β4 subunits onto the α and δ subunits respectively of the *Torpedo marmorata* cryoelectron microscopic structure \(^\text{152}\); PDB 2BG9) using the SWISS-MODEL web tool (http://swissmodel.expasy.org/). Structures were visualized and displays generated using Chimera 1.6.2 (http://www.cgl.ucsf.edu/chimera). The subunits are shown as chain trace (α3 red, β4 cyan) with the ACh-binding loops highlighted (α3 yellow, β4 blue). Variants are shown as stick and ball: α3(R37) (red), β4(K57) - dark blue, β4(T91) - orange, β4(R136) - pink, β4(S140) - purple, β4(G296) - dark green, β4(R349) - in unstructured region (not shown), β4(T375) - in unstructured region, β4(M456) - dark green, β4(M467) - light green. None of the variants are in the transmitter-binding site, nor are they at regions likely to be points of interaction between subunits. β4(G296) is located in the 3rd and β4(M467) in the 4th membrane spanning region, but none are located in the channel/gate regions of the 2nd membrane spanning region.
Figure 4.41. Responses to acetylcholine and nicotine from cells expressing wild-type α3 and β4 subunits. Panel A shows data from one cell for responses to ACh, and panel B responses to nicotine (from a different cell). The traces in red are response to the highest concentration; note that the responses decline more rapidly and show a "tail" of increased response when the agonist is removed. This is characteristic of channel block by agonist. On the right, the normalized concentration-response curves are shown; red triangles show data from the traces on the left, while blue circles show the mean (± SE) data for 20 cells (ACh) or 9 cells (nicotine). The lines show predictions of the Hill equation:

\[
\text{response} = \text{Amp} \times \left(\frac{[\text{agonist}]}{\text{EC}_{50}}\right)^{n_{\text{Hill}}} / \left(1 + \left(\frac{[\text{agonist}]}{\text{EC}_{50}}\right)^{n_{\text{Hill}}} \right)
\]

where Amp is the maximal response, EC_{50} is the concentration producing a half-maximal effect, and n_{Hill} is the Hill coefficient. For individual cell data the values used were the best-fitting values for that cell while for the mean data the values were the mean values for all fits to data obtained with that agonist. Data for nicotine were normalized to the maximal response to ACh.
Figure 4.42. Parameters normalized to that for wild-type, for receptors containing combinations of α3(R37H) with β4 variants. The panels show the surface expression, then the relative efficacy for nicotine, the EC_{50} for nicotine and the EC_{50} for ACh. Each pair of bars shows data for α3 or α3(R37H) expressed with the specified β4 subunit. The symbols on the right show significance of the difference to α3β4, first for α3 with the β4 subunit then for α3(R37H) (ns: >0.05, * <0.05; ** < 0.01; *** < 0.001). (Data are mean±SE).
Figure 4.43. Parameters normalized to that for wild-type, for receptors containing combinations of β4(M467V) with other variants, all expressed with α3. Variants were combined on the same subunit (e.g. β4(M467V)+S140G), or expressed as a 1:1 mixture of subunits (e.g. β4(M467V)&β4(S140G)). The figure is in the same format as previous figures.
Figure 4.44. Surface expression for \( \beta_4 \) variants expressed with the \( \alpha_3 \) or the \( \alpha_4 \) subunit. Each pair of bars shows data for the specified \( \beta_4 \) subunit expressed with either the \( \alpha_3 \) or \( \alpha_4 \) subunit. The symbols on the right show significance of the difference to expression of wild-type \( \beta_4 \), first for \( \alpha_3 \) then for \( \alpha_4 \) (ns: >0.05, * <0.05; ** < 0.01; *** < 0.001). (Data are mean + SE).
Figure 4.45. Physiological parameters normalized for β4 variants normalized to values for wild-type β4. Each pair of bars shows data for the indicated variant expressed with either α3 or α4 subunits. The symbols on the right show significance of the difference to parameters for wild-type β4, first for α3 then for α4 (ND not determined; ns: >0.05; * <0.05; ** < 0.01; *** < 0.001). (Data are mean± SE).
Figure 4.46. Scatter plot of expression of β4 variants with the α4 subunit against expression with the α3 subunit. Each point shows the mean ± SE, while the solid point shows the value for wild-type β4. The dashed line shows the line of equality, and the solid line shows the regression relationship (slope = 1.36, different from a slope of 0 with $P = 0.01$). The relationship is consistent with the idea that the variants do not have a selective effect on expression with a specific α subunit, but rather affect maturation, general ability to be assembled or trafficking to the surface.
Figure 4.47. Scatter plots of the average maximal response to ACh for expression of β4 variants against the surface expression. The left panel shows data for variants expressed with the α3 subunit, the right panel for expression with the α4 subunit. Each point shows the mean ± SE, while the solid point shows the value for wild-type β4. In each panel the dashed lines show the line of equality, and the solid lines show the regression relationship. For expression with α3 the slope = 0.54 that differs from a slope of 0 with P = 0.01. For expression with α4 the slope = 0.17 that does not differ from a slope of 0 (P = 0.5). The maximal response was determined with cells selected for high surface expression while the surface ELISA measures the average surface expression per cell. Accordingly, it might be expected that variants with low expression would have a larger relative response compared to wild-type than surface ELISA.
<table>
<thead>
<tr>
<th>Receptor</th>
<th>ACh EC$_{50}$ (μM)</th>
<th>Nic EC$_{50}$ (μM)</th>
<th>Nicotine efficacy</th>
<th>Imax (pA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a4&amp;b4</td>
<td>11 ± 2 (6) ns</td>
<td>0.99 ± 0.24 (5)</td>
<td>0.86 ± 0.07 (5) ns</td>
<td>1087 ± 217 (11)</td>
</tr>
<tr>
<td>a4&amp;b4(K57E)</td>
<td>8 ± 2 (4) ns</td>
<td>0.72 ± 0.07 (10) ns</td>
<td>0.95 ± 0.10 (10) ns</td>
<td>879 ± 137 (15) ns</td>
</tr>
<tr>
<td>a4&amp;b4(T911)</td>
<td>13 ± 6 (4) ns</td>
<td>0.56 ± 0.09 (5) ns</td>
<td>1.07 ± 0.04 (5) ns</td>
<td>1424 ± 198 (10) ns</td>
</tr>
<tr>
<td>a4&amp;b4(R136W)</td>
<td>12 ± 4 (5) ns</td>
<td>0.42 ± 0.09 (5) ns</td>
<td>0.86 ± 0.09 (5) ns</td>
<td>944 ± 225 (10) ns</td>
</tr>
<tr>
<td>a4&amp;b4(N138A)</td>
<td>12 ± 2 (4) ns</td>
<td>2.86 ± 0.37 (10) *</td>
<td>0.78 ± 0.05 (10) ns</td>
<td>749 ± 143 (14) ns</td>
</tr>
<tr>
<td>a4&amp;b4(S140G)</td>
<td>21 ± 2 (10) ns</td>
<td>1.14 ± 0.19 (10) ns</td>
<td>0.93 ± 0.09 (10) ns</td>
<td>956 ± 110 (31) ns</td>
</tr>
<tr>
<td>a4&amp;b4(T3751)</td>
<td>13 ± 4 (4) ns</td>
<td>0.21 ± 0.03 (5) ns</td>
<td>1.05 ± 0.04 (5) ns</td>
<td>1331 ± 198 (9) ns</td>
</tr>
<tr>
<td>a4&amp;b4(M467V)</td>
<td>21 ± 4 (5) ns</td>
<td>0.63 ± 0.09 (5) ns</td>
<td>0.77 ± 0.11 (5) ns</td>
<td>684 ± 121 (10) ns</td>
</tr>
</tbody>
</table>

Table 4.1. Physiological results for variants expressed with the α4 subunit. The first column gives the combination of subunits expressed. The next 4 columns give data as mean ± SE (number of cells). The last entry is the significance of the difference to wild-type β4 (one-way ANOVA with Dunnett’s correction; ns > 0.05; * < 0.05; ** < 0.01; *** < 0.001). The EC$_{50}$ value gives the concentration of agonist producing a half-maximal response, "Nicotine efficacy" is the ratio of the maximal nicotine response to the maximal ACh response, Imax gives the negative of the predicted maximal response of a cell to ACh.
<table>
<thead>
<tr>
<th>Receptor</th>
<th>ACh EC$_{50}$ (µM)</th>
<th>Nic EC$_{50}$ (µM)</th>
<th>Nicotine efficacy</th>
<th>I$_{max}$ (-pA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a3&amp;b4</td>
<td>146 ± 34 (20)</td>
<td>22 ± 4 (9)</td>
<td>0.73 ± 0.03 (17)</td>
<td>1541 ± 165 (37) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R18C)</td>
<td>410 ± 100 (10)**</td>
<td>29 ± 7 (5) ns</td>
<td>0.63 ± 0.06 (5) ns</td>
<td>3488 ± 558 (15)**</td>
</tr>
<tr>
<td>a3&amp;b4(K57E)</td>
<td>110 ± 23 (5) ns</td>
<td>62 ± 11 (9)**</td>
<td>0.79 ± 0.03 (18) ns</td>
<td>417 ± 43 (22)**</td>
</tr>
<tr>
<td>a3&amp;b4(T91I)</td>
<td>246 ± 51 (7) ns</td>
<td>25 ± 6 (8) ns</td>
<td>0.77 ± 0.04 (15) ns</td>
<td>837 ± 135 (27) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R136Q)</td>
<td>127 ± 29 (6) ns</td>
<td>56 ± 10 (5) ns</td>
<td>0.69 ± 0.05 (5) ns</td>
<td>1005 ± 126 (11) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R136W)</td>
<td>72 ± 6 (22) ns</td>
<td>28 ± 4 (10) ns</td>
<td>0.84 ± 0.06 (22) ns</td>
<td>1362 ± 175 (44) ns</td>
</tr>
<tr>
<td>a3&amp;b4(N138A)</td>
<td>116 ± 20 (2) ns</td>
<td>29 ± 5 (5) ns</td>
<td>0.54 ± 0.03 (5) ns</td>
<td>420 ± 93 (9) ns</td>
</tr>
<tr>
<td>a3&amp;b4(S140G)</td>
<td>121 ± 20 (12) ns</td>
<td>48 ± 4 (18) ns</td>
<td>0.74 ± 0.03 (18) ns</td>
<td>1113 ± 170 (30) ns</td>
</tr>
<tr>
<td>a3&amp;b4(S140T)</td>
<td>99 ± 13 (6) ns</td>
<td>13 ± 2 (4) ns</td>
<td>0.75 ± 0.06 (4) ns</td>
<td>2314 ± 619 (10) ns</td>
</tr>
<tr>
<td>a3&amp;b4(G296S)</td>
<td>57 ± 38 (2) ns</td>
<td>19 ± 3 (6) ns</td>
<td>0.74 ± 0.05 (6) ns</td>
<td>1558 ± 315 (11) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R349C)</td>
<td>101 ± 23 (5) ns</td>
<td>26 ± 6 (4) ns</td>
<td>0.79 ± 0.07 (10) ns</td>
<td>216 ± 42 (15)**</td>
</tr>
<tr>
<td>a3&amp;b4(T375I)</td>
<td>227 ± 77 (10) ns</td>
<td>33 ± 12 (7) ns</td>
<td>0.81 ± 0.04 (17) ns</td>
<td>1266 ± 108 (30) ns</td>
</tr>
<tr>
<td>a3&amp;b4(A435T)</td>
<td>131 ± 30 (6) ns</td>
<td>43 ± 5 (5) ns</td>
<td>0.59 ± 0.06 (5) ns</td>
<td>1999 ± 534 (11) ns</td>
</tr>
<tr>
<td>a3&amp;b4(A435V)</td>
<td>119 ± 26 (6) ns</td>
<td>18 ± 2 (5) ns</td>
<td>0.76 ± 0.02 (5) ns</td>
<td>1505 ± 359 (11) ns</td>
</tr>
<tr>
<td>a3&amp;b4(M456V)</td>
<td>173 ± 26 (2) ns</td>
<td>69 ± 23 (5)**</td>
<td>0.77 ± 0.07 (5) ns</td>
<td>1077 ± 169 (10) ns</td>
</tr>
<tr>
<td>a3&amp;b4(M467V)</td>
<td>124 ± 20 (12) ns</td>
<td>38 ± 6 (12) ns</td>
<td>0.86 ± 0.07 (18) ns</td>
<td>1289 ± 220 (33) ns</td>
</tr>
<tr>
<td>a3&amp;b4(M467V+R136W)</td>
<td>54 ± 7 (16) ns</td>
<td>30 ± 5 (17) ns</td>
<td>0.61 ± 0.02 (17) ns</td>
<td>1731 ± 241 (33) ns</td>
</tr>
<tr>
<td>a3&amp;b4(M467V+S140G)</td>
<td>163 ± 37 (12) ns</td>
<td>18 ± 3 (7) ns</td>
<td>0.63 ± 0.06 (7) ns</td>
<td>1237 ± 212 (19) ns</td>
</tr>
<tr>
<td>a3MIXwt(S140G)</td>
<td>71 ± 13 (7) ns</td>
<td>30 ± 4 (8) ns</td>
<td>0.76 ± 0.03 (8) ns</td>
<td>2135 ± 304 (15) ns</td>
</tr>
<tr>
<td>a3MIX(M467V)(S140G)</td>
<td>58 ± 7 (6) ns</td>
<td>43 ± 6 (10) ns</td>
<td>0.77 ± 0.05 (14) ns</td>
<td>2423 ± 439 (6) ns</td>
</tr>
<tr>
<td>a3MIX(M467V)(R136W)</td>
<td>140 ± 21 (14) ns</td>
<td>26 ± 2 (17) ns</td>
<td>0.73 ± 0.03 (20) ns</td>
<td>1981 ± 306 (28) ns</td>
</tr>
<tr>
<td>a3(R37H)&amp;b4</td>
<td>86 ± 30 (3) ns</td>
<td>7 ± 2 (4) ns</td>
<td>0.98 ± 0.10 (7) *</td>
<td>1027 ± 234 (13) ns</td>
</tr>
<tr>
<td>a3(R37H)&amp;b4(T91I)</td>
<td>95 ± 16 (12) ns</td>
<td>23 ± 3 (10) ns</td>
<td>0.95 ± 0.10 (10) *</td>
<td>1509 ± 273 (22) ns</td>
</tr>
<tr>
<td>a3(R37H)&amp;b4(T375I)</td>
<td>71 ± 12 (4) ns</td>
<td>20 ± 2 (5) ns</td>
<td>0.88 ± 0.03 (10) ns</td>
<td>1779 ± 304 (14) ns</td>
</tr>
</tbody>
</table>

Table 4.2. Physiological results for variants expressed with the a3 subunit. The first column gives the combination of subunits expressed. The next 4 columns give data as mean ± SE (number of cells). The last entry is the significance of the difference to wild-type b4 (one-way ANOVA with Dunnett's correction; ns > 0.05; * < 0.05; ** < 0.01; *** < 0.001). The EC$_{50}$ value gives the concentration of agonist producing a half-maximal response, "Nicotine efficacy" is the ratio of the maximal nicotine response to the maximal ACh response, I$_{max}$ gives the negative of the predicted maximal response of a cell to ACh.
<table>
<thead>
<tr>
<th>Receptor</th>
<th>Relative expression</th>
<th>Relative cell protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>a4&amp;b4</td>
<td>1.00 ± 0.00 (24)</td>
<td>1.10 ± 0.07 (24) NS</td>
</tr>
<tr>
<td>a4&amp;b4(K57E)</td>
<td>0.03 ± 0.03 (3) ***</td>
<td>1.17 ± 0.12 (3) NS</td>
</tr>
<tr>
<td>a4&amp;b4(T91I)</td>
<td>0.31 ± 0.05 (2) **</td>
<td>0.94 ± 0.06 (2) NS</td>
</tr>
<tr>
<td>a4&amp;b4(R136Q)</td>
<td>0.88 ± 0.25 (3) NS</td>
<td>0.85 ± 0.12 (3) NS</td>
</tr>
<tr>
<td>a4&amp;b4(R136W)</td>
<td>0.53 ± 0.18 (3) *</td>
<td>1.04 ± 0.28 (3) NS</td>
</tr>
<tr>
<td>a4&amp;b4(N138A)</td>
<td>0.04 ± 0.08 (3) ***</td>
<td>1.07 ± 0.30 (3)</td>
</tr>
<tr>
<td>a4&amp;b4(S140G)</td>
<td>-0.13 ± 0.22 (2) ***</td>
<td>0.93 ± 0.34 (2) NS</td>
</tr>
<tr>
<td>a4&amp;b4(G296S)</td>
<td>2.26 ± 0.27 (6) ***</td>
<td>1.07 ± 0.13 (6) NS</td>
</tr>
<tr>
<td>a4&amp;b4(R349C)</td>
<td>0.14 ± 0.07 (4) ***</td>
<td>1.20 ± 0.18 (4) NS</td>
</tr>
<tr>
<td>a4&amp;b4(T375I)</td>
<td>0.68 ± 0.04 (2) NS</td>
<td>0.87 ± 0.06 (2) NS</td>
</tr>
<tr>
<td>a4&amp;b4(M456V)</td>
<td>0.60 ± 0.09 (4) *</td>
<td>1.03 ± 0.01 (4) NS</td>
</tr>
<tr>
<td>a4&amp;b4(M467V)</td>
<td>0.60 ± 0.24 (4) *</td>
<td>1.35 ± 0.36 (4) NS</td>
</tr>
<tr>
<td>a4</td>
<td>0.07 ± 0.04 (9) ***</td>
<td>1.00 ± 0.13 (9) NS</td>
</tr>
</tbody>
</table>

Table 4.3. Surface expression for variants expressed with the α4 subunit. The first column gives the combination of subunits expressed. The next 2 columns give data as mean ± SE (number of cells). The last entry is the significance of the difference to wild-type β4 (one-way ANOVA with Dunnett's correction; ns > 0.05; * < 0.05; ** < 0.01; *** < 0.001). Relative expression gives the surface expression relative to wild-type β4 and the last entry is the significance of the difference to wild-type β4 (one-way ANOVA with Dunnett's correction; ns > 0.05; * < 0.05; ** < 0.01; *** < 0.001). Relative cell protein gives the total protein per well relative to pcDNA3 and the last entry is the significance of the difference to pcDNA3 (one-way ANOVA with Dunnett's correction; ns > 0.05; * < 0.05; ** < 0.01; *** < 0.001).
<table>
<thead>
<tr>
<th>Receptor</th>
<th>Relative expression</th>
<th>Relative cell protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>a3&amp;b4</td>
<td>1.00 ± 0.00 (60)</td>
<td>0.88 ± 0.03 (60) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R18C)</td>
<td>1.45 ± 0.30 (3) **</td>
<td>0.92 ± 0.18 (3) ns</td>
</tr>
<tr>
<td>a3&amp;b4(K57E)</td>
<td>-0.06 ± 0.14 (6) ***</td>
<td>1.10 ± 0.08 (6) ns</td>
</tr>
<tr>
<td>a3&amp;b4(T911)</td>
<td>0.27 ± 0.06 (5) ***</td>
<td>0.90 ± 0.07 (5) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R136Q)</td>
<td>0.16 ± 0.16 (8) ***</td>
<td>1.01 ± 0.07 (8) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R136W)</td>
<td>0.32 ± 0.09 (9) ***</td>
<td>0.99 ± 0.06 (9) ns</td>
</tr>
<tr>
<td>a3&amp;b4(N138A)</td>
<td>0.16 ± 0.08 (7) ***</td>
<td>0.73 ± 0.10 (7) ns</td>
</tr>
<tr>
<td>a3&amp;b4(S140G)</td>
<td>-0.04 ± 0.05 (12) ***</td>
<td>0.99 ± 0.10 (12) ns</td>
</tr>
<tr>
<td>a3&amp;b4(S140T)</td>
<td>0.88 ± 0.15 (4) ns</td>
<td>0.70 ± 0.04 (4) ns</td>
</tr>
<tr>
<td>a3&amp;b4(N138A+S140G)</td>
<td>0.02 ± 0.01 (3) ***</td>
<td>0.94 ± 0.07 (3) ns</td>
</tr>
<tr>
<td>a3&amp;b4(G296S)</td>
<td>0.76 ± 0.31 (5) ns</td>
<td>1.37 ± 0.27 (5) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R349C)</td>
<td>0.24 ± 0.05 (7) ***</td>
<td>1.29 ± 0.50 (7) ns</td>
</tr>
<tr>
<td>a3&amp;b4(T375I)</td>
<td>0.40 ± 0.08 (5) ***</td>
<td>0.95 ± 0.12 (5) ns</td>
</tr>
<tr>
<td>a3&amp;b4(R421X)</td>
<td>0.04 ± 0.01 (2) ***</td>
<td>0.73 ± 0.07 (2) ns</td>
</tr>
<tr>
<td>a3&amp;b4(A435T)</td>
<td>0.73 ± 0.11 (3) ns</td>
<td>0.83 ± 0.08 (3) ns</td>
</tr>
<tr>
<td>a3&amp;b4(A435V)</td>
<td>1.12 ± 0.12 (3) ns</td>
<td>0.77 ± 0.06 (3) ns</td>
</tr>
<tr>
<td>a3&amp;b4(M456V)</td>
<td>0.22 ± 0.11 (7) ***</td>
<td>1.08 ± 0.11 (7) ns</td>
</tr>
<tr>
<td>a3&amp;b4(M467V)</td>
<td>0.73 ± 0.22 (7) *</td>
<td>1.16 ± 0.23 (7) ns</td>
</tr>
<tr>
<td>a3&amp;b4(M467V+R136W)</td>
<td>0.61 ± 0.02 (2) ns</td>
<td>0.77 ± 0.02 (2) ns</td>
</tr>
<tr>
<td>a3&amp;b4(M467V+S140G)</td>
<td>0.17 ± 0.02 (2) ***</td>
<td>0.66 ± 0.01 (2) ns</td>
</tr>
<tr>
<td>a3&amp;MXb4(M467V)+b4(S140G)</td>
<td>0.75 ± 0.18 (4) ns</td>
<td>0.86 ± 0.28 (4) ns</td>
</tr>
<tr>
<td>a3&amp;MXb4(M467V)+b4(R136W)</td>
<td>0.52 ± 0.02 (4) ***</td>
<td>0.69 ± 0.31 (4) ns</td>
</tr>
<tr>
<td>a3(R37H)&amp;b4</td>
<td>0.06 ± 0.02 (5) ***</td>
<td>1.24 ± 0.16 (5) ns</td>
</tr>
<tr>
<td>a3(R37H)&amp;b4(T911)</td>
<td>0.03 ± 0.02 (5) ***</td>
<td>1.31 ± 0.26 (5) ns</td>
</tr>
<tr>
<td>a3(R37H)&amp;b4(T375I)</td>
<td>0.07 ± 0.03 (5) ***</td>
<td>1.28 ± 0.15 (5) ns</td>
</tr>
<tr>
<td>a3</td>
<td>-0.01 ± 0.00 (2) ***</td>
<td>1.10 ± 0.09 (2) ns</td>
</tr>
</tbody>
</table>

Table 4.4. Surface expression for variants expressed with the a3 subunit. The first column gives the combination of subunits expressed. The next 2 columns give data as mean ± SE (number of cells). The last entry is the significance of the difference to wild-type β4 (one-way ANOVA with Dunnett's correction; ns > 0.05; * < 0.05; ** < 0.01; *** < 0.001). Relative expression gives the surface expression relative to wild-type β4 and the last entry is the significance of the difference to wild-type β4 (one-way ANOVA with Dunnett's correction; ns > 0.05; * < 0.05; ** < 0.01; *** < 0.001). Relative cell protein gives the total protein per well relative to pcDNA3 and the last entry is the significance of the difference to pcDNA3 (one-way ANOVA with Dunnett's correction; ns > 0.05; * < 0.05; ** < 0.01; *** < 0.001).
Figure 4.48. Normalized cigarettes per day among carriers of CHRNA4 missense variants. Residuals of log transformed cigarettes per day (CPD) after correcting for age and sex were compared between individuals carrying at least one missense variant in \textit{CHRNA4} and those not carrying any missense variants in \textit{CHRNA4} using linear regression in R.
Figure 4.49. Normalized cigarettes per day among carriers of CHRNA4 missense variants at conserved (vertebrate PhyloP score >2) sites. Residuals of log transformed cigarettes per day (CPD) after correcting for age and sex were compared between individuals carrying at least one missense variant in CHRNA4 at a conserved site and those not carrying missense variants at conserved sites using linear regression in R.
A \[ r^2 = 0.029 \ ; \beta = -0.006 \ ; p\text{-value} = 6 \times 10^{-6} \]

Figure 4.50. Correlation between Acetylcholine EC50 and CPD. A) The distribution of log(CPD) after correcting for age and sex for each *CHRNB4* variant’s estimated acetylcholine EC\(_{50}\). Red line is the estimated linear relationship. B) Box plots of log(CPD) after correcting for age and sex for each variant or group of variants (i.e. individuals with two variants) ordered by acetylcholine EC50 value.
$r^2 = 0.0008 ; \beta = 0.003 ; p\text{-value} = 0.43$

**Figure 4.51.** Correlation between Nicotine EC$_{50}$ and CPD. A) The distribution of log(CPD) after correcting for age and sex for each $CHRN B4$ variant’s estimated nicotine EC$_{50}$. Red line is the estimated linear relationship. B) Box plots of log(CPD) after correcting for age and sex for each variant or group of variants (i.e. individuals with two variants) ordered by nicotine EC$_{50}$ value.
Figure 4.52. Correlation between Nicotine efficacy and CPD. A) The distribution of log(CPD) after correcting for age and sex for each CHRNA4 variant’s estimated nicotine efficacy. Red line is the estimated linear relationship. B) Box plots of log(CPD) after correcting for age and sex for each variant or group of variants (i.e. individuals with two variants) ordered by nicotine efficacy value.
Figure 4.53. Correlation between Cell-Surface Expression of \textit{CHRNB4} variants when expressed with \textit{CHRNA3} and CPD. A) The distribution of log(CPD) after correcting for age and sex for each \textit{CHRNB4} variant’s estimated cell-surface expression. Red line is the estimated linear relationship. B) Box plots of log(CPD) after correcting for age and sex for each variant or group of variants (i.e. individuals with two variants) ordered by cell-surface expression value.
Figure 4.54. Correlation between acetylcholine maximal response and CPD. A) The distribution of log(CPD) after correcting for age and sex for each $CHRNB4$ variant's estimated acetylcholine maximal response. Red line is the estimated linear relationship. B) Box plots of log(CPD) after correcting for age and sex for each variant or group of variants (i.e. individuals with two variants) ordered by acetylcholine maximal response value.

A  $r^2 = 0.003 \ ; \beta = -1.731 \ ; \text{p-value} = 0.11$
Chapter 5. Conclusions and Future Directions

*State of substance dependence research prior to this work*

Despite the efforts of many candidate gene studies and several large meta-analyses consisting of >40,000 individuals, very few genes that contribute to the risk of nicotine dependence or cigarette consumption have been identified. Further, the genes identified have been excellent candidate genes. The two classes of genes that have been identified as contributors to nicotine dependence risk are nicotinic acetylcholine receptors, *CHRNA5/B4/A3* on chromosome 15 and *CHRNA6/B3* on chromosome 8, and *CYP2A6*, the major nicotine metabolizing enzyme, the only non-nicotinic receptor gene to be significantly associated with the number of cigarettes smoked per day (CPD) at the genome-wide level. A few sequencing studies had investigated the role of rare variants in select nicotinic acetylcholine receptor genes prior to the commencement of this work. These studies were limited to cohorts of European Americans and only limited numbers of individuals were sequenced. The only survey of common variation and nicotine dependence, as opposed to CPD, was in a tandem candidate gene and genome-wide survey that examined 3718 SNPs in 348 candidate genes in conjunction with a pooled genome-wide association study of >2.4 million variants followed by individual genotyping of ~40,000 variants. These studies were the first to report the association between SNPs near *CHRNA5/B4/A3* and *CHRNA6/B3* with nicotine dependence. Like the state of nicotine dependence research prior to this work, alcohol and cocaine dependence research had identified only a small number of genes and variants that were consistently associated with either alcohol or cocaine dependence. Among these are
several weakly associated variants in a number of gamma-amino butyric acid receptors (GABRA), alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), the mu opioid receptor (OPRM1) and bitter taste receptors (TAS) genes (see 63 for review). The first genome-wide significant association for alcohol dependence was with a missense variant (rs1229984; R48H) in the ADHIB gene 153. Our group recently demonstrated the only other genome-wide significant association with alcohol dependence, a bin of common SNPs tagged by rs12912251 near C15orf53 64. The studies of alcohol dependence were not as well powered as the studies of cigarette consumption. At most, these alcohol dependence studies were performed with a few thousand individuals while the meta-analyses of cigarette consumption were performed with 40-70,000 individuals. No variant has ever been shown to be genome-wide significantly associated with cocaine dependence, however. Of course, few studies have been sufficiently powered to detect effects of the size expected (odds-ratios of 1.1-1.5) and no genome-wide association studies of cocaine dependence have been performed. Additionally, no sequencing studies had been done prior to this work for either alcohol or cocaine dependence.

Much work had been done prior to the commencement of this work on functionally characterizing the nicotinic acetylcholine receptor genes. For instance, mice lacking various nicotinic receptor genes have been shown to have altered behavioral response to nicotine 77,90,154. Little work had been done, however, describing the effects of variants segregating within human populations within these nicotinic acetylcholine receptor genes. Two such studies characterized a handful of variants in CHRNA3, CHRNA4 and CHRNB4, but did so in xenopus oocytes and studies of CHRNB4 variants were done in combination with CHRNA4 rather than CHRNA3 155,156. Work from our
group has shown striking behavioral effects of knocking in the D398N allele into the mouse, but this work is yet unpublished. Additionally, no studies have been done wherein functional information about rare variants has been integrated into association analyses. Some studies have subset, as we have, on variants with higher prior probability of being functional (i.e. missense, nonsense, splice-site), but none to our knowledge have weighted genotypes based on the results from a functional assay.

**Dissertation specific aims**

As very little of the variance in nicotine dependence risk, and the variance in nearly all complex traits, can be explained by common SNPs, we began investigating the potential role of rare genetic variants in risk for this trait. We undertook pooled sequencing of the coding regions and flanking sequence of the *CHRNA5*, *CHRNA3*, *CHRNB4*, *CHRNA6* and *CHRNB3* genes in 710 African Americans (461 nicotine dependent (ND) cases and 249 smokers with no symptoms of dependence (controls)) and 2055 European Americans (1062 ND cases and 993 controls) from the Collaborative Study of the Genetics of Nicotine Dependence (COGEND). We then asked whether the proportion of individuals harboring at least one rare non-synonymous variant at the subset of amino-acid positions conserved across vertebrate species was different in nicotine dependent smokers compared to non-nicotine dependent smokers. We demonstrated a reduced risk for nicotine dependence among carriers of missense variants at conserved residues in *CHRNB4* in African Americans and European Americans.

We then aimed to determine whether rare genetic variation in these genes influence risk for developing alcohol or cocaine dependence, two conditions highly
comorbid with nicotine dependence. We again undertook pooled sequencing of the coding regions and flanking sequence of the CHRNA5, CHRNA3, CHRNA6 and CHRNA3 genes in 287 African Americans (147 DSM-IV alcohol dependent cases and 140 controls) and 1028 European Americans (480 DSM-IV alcohol dependent cases and 548 controls) individuals from the Collaborative Study of the Genetics of Alcoholism (COGA). Of these individuals, there were 98 (66%) African American DSM-IV alcohol dependent cases who were also DSM-IV cocaine dependent and 184 (38%) European American DSM-IV alcohol dependent cases who were also DSM-IV cocaine dependent. We then performed follow-up genotyping of all identified non-synonymous variants in all available members of the families from which the sequenced individuals came, totaling 2504 African Americans and 7318 European Americans. For European Americans, we found increased DSM-IV cocaine dependence symptoms and increased DSM-IV alcohol dependence symptoms among carriers of missense variants in CHRNA3. For African Americans, we find decreased cocaine dependence symptoms among carriers of missense variants in CHRNA3. These two genes had not previously been implicated in alcohol or cocaine dependence specifically. The gene cluster containing the CHRNA5, CHRNA4 and CHRNA3 genes was found to be associated with alcohol dependence and cocaine dependence, but the linkage disequilibrium bins associated with these phenotypes contain SNPs that alter CHRNA5 brain mRNA expression (bin tagged by rs588765) and alter protein function (rs16969968, CHRNA5 D398N), neither implicating CHRNA3 or CHRNA4 directly.

Finally, we sought to determine the functional impact of rare variants in CHRNA4 harbored by individuals in these two cohorts. All variants observed in at least one
individual in the COGEND cohort in \textit{CHRNA3} (14 variants) were investigated. One variant in \textit{CHRNA3} (R37H) was also functionally tested as it is in high linkage disequilibrium with another rare missense variant (T91I) in \textit{CHRNA3}. We found several variants that alter cellular response either to nicotine or acetylcholine as well as many variants that alter cell-surface protein expression as measured by cell-surface ELISA without altering either \textit{CHRNA4} mRNA or total β4 protein as measured by western blot. Further, when we integrate these \textit{in vitro} findings into a model of association with nicotine dependence related traits, we were able to substantially improve the association in both African Americans and European Americans. This work strongly suggesting that the success of future association analyses in these and potentially many other genes across the genome, may depend greatly on functional assessment of observed genetic variation. To my knowledge, no other group has attempted to integrate the results from function assays directly into association analyses. Many groups, ourselves included, have performed tests using only variants fitting some number of criteria, for instance variants that occur in functional domains, that perform above a certain threshold on a functional assay or occur at evolutionarily conserved sites within the genome, but no group has integrated quantitative data regarding the functional effect of variants composing the collapsed genotypes using in their analyses.

These findings support the assertion that variants in or near these nicotinic receptor genes are capable of pleiotropic effects. Previous work from our group has shown that the common missense variant in \textit{CHRNA5} (D398N, rs16969968) affects both nicotine and cocaine dependence and that a group of SNPs affecting \textit{CHRNA5} mRNA expression in the human frontal cortex contribute to alcohol dependence risk and nicotine
dependence risk\textsuperscript{45,139}. In both the case of \textit{CHRNA5} and \textit{CHRNA6-CHRNB3}, the allele that increases risk for nicotine dependence lowers risk for cocaine dependence or alcohol dependence, a fact that makes comorbidity an unlikely cause of the pleiotropy observed. The results from this dissertation suggest that rare variants in nicotinic receptors can contribute to multiple substance dependence phenotypes. We do not, however, observe a rare variant association at any gene and more than one phenotype independently of the other phenotypes. We observe that rare variants in \textit{CHRNB4} affect nicotine dependence and only nicotine dependence and that rare variants in \textit{CHRNB3} affect alcohol and cocaine dependence. It is possible that these findings are the result of having only minimal power to detect such rare variant associations, even given the relatively large numbers of individuals sequenced as part of this thesis.

Overall, this dissertation has identified one novel gene for nicotine dependence and two novel genes for cocaine/alcohol dependence, that the variants underlying the association of \textit{CHRNB4} with nicotine dependence have a significant effect on receptor function \textit{in vitro} and that the magnitude and direction of the functional effect of \textit{CHRNB4} variants on acetylcholine EC\textsubscript{50} can be used to improve the observed association with nicotine dependence. These are some of the first results implicating rare variants in the etiology of human behaviors (nicotine, alcohol and cocaine dependence) and some of the first results to identify rare variants in genes that do not cause a more severe, Mendelian form of the complex trait under investigation. For instance, \textit{PCSK9}, one of the first examples of a gene containing rare variants contributing to a complex trait, encodes a protease that reduces low-density lipoprotein (LDL) receptor levels. Mutations in \textit{PCSK9} had been previously show to cause dominant forms of hypercholesterolemia, so in many
ways the association of rare variants in this gene with cholesterol levels was unsurprising. Our results suggest that the identification of associations with either common or rare variants with truly complex traits require exceedingly large numbers of individuals (we sequenced hundreds to thousands of individuals and performed genotyping of observed variants in thousands of individuals), will require that those individuals be well characterized phenotypically and will likely also require functional characterization of observed variants to filter association signal from background noise, i.e. the plethora of neutral alleles extant in human populations.

**Future directions**

While one novel gene was found to be associated with nicotine dependence and two genes were identified that contribute to risk for alcohol and/or cocaine dependence as the result of this work, additional effort will be needed to further validate these findings and better understand the contribution of these genes to the phenotypes they affect. This will likely be done by sequencing of larger numbers of individuals to look for associations at these loci with substance dependence phenotypes and to functionally characterize identified associated variants. Further work is necessary to find as of yet undiscovered genes and variants that contribute to substance dependence risk. Additionally, this work strongly suggests that some variants that contribute to risk of one substance dependence may contribute to multiple substance dependences and that further investigation into the effect of genetic variants on comorbid phenotypes is warranted. The fact that multiple variants in *CHRNB3* contribute to both alcohol and cocaine dependence and the fact that these variants occur in a gene that harbors common genetic variants associated with both
nicotine dependence and cigarette consumption, strongly implicates these nicotinic receptor genes specifically in determining one’s risk of becoming dependent on multiple substances. In the future, designing a study such that the effects of variants on specific drugs independently of one another would be ideal. For instance, having an entire sample composed of individuals with no history of drug use or exposure to any drug other than the drug under study would ensure that any association observed is due to that specific drug. Then if you saw the same variant associated with another drug in a cohort designed similarly, i.e. dependent individuals are dependent only on the drug under investigation, then one could more conclusively say that the variant has pleiotropic effects on various substance dependence phenotypes. A design in which cases are only dependent on one drug, especially if that drug is illegal, may be impossible, however. Currently, studies have been designed to maximize power to detect associations between genetic variants and a specific drug but have not excluded individuals dependent on other drugs. This design has made it difficult to determine the dependence the variant is truly affecting. At the very least, improved statistical models are required to de-convolute the complex interactions between these drugs, environmental factors and genetic contributors.

An expanded effort to identify variants in the genes studied here and to identify variants and associations in an unbiased manner, via whole-exome or whole-genome sequencing of affected and unaffected individuals, will be necessary to improve our understanding of how genetic factors contribute to substance dependence risk. Our group has begun sequencing the genomic regions containing the nicotinic receptor genes, \textit{CYP2A6} and a number of other nicotine metabolizing enzymes in a large number of individuals from COGEND. We have obtained African American samples in a similar
ascertainment scheme to that used for COGEND as part of the African American Study of Nicotine Dependence (AAND) and have begun sequencing this cohort as well. This effort will comprehensively evaluate the contribution of both common and rare variants in these regions with respect to their contribution to nicotine dependence risk. While most of these individuals were sequenced as part of this dissertation, several hundred were merely genotyped for variants observed in sequenced individuals. It is possible, then, that the added variants discovered in these newly sequenced individuals will give the rare variant associations the power needed to observe significant associations between some of these genes and nicotine dependence. Of course, the next important step in determining the genetic underpinnings of nicotine dependence is to perform either exome or whole-genome sequencing in this cohort, or ideally an expanded cohort, to test for associations between rare variants in each gene across the genome and risk for nicotine dependence. When sequencing becomes sufficiently inexpensive, this study is an inevitability. Given the expensive nature of whole-genome sequencing and even exome sequencing at this point in time, paired with the fact that very large numbers of individuals are necessary to obtain sufficient power to observe significant genetic associations either with common variants, collapsed rare variants or rare variants using statistical methods like SKAT or the c-alpha test, it is likely most efficient to perform targeted sequencing of a large number of individuals. If one has sufficient funds, however, performing exome-sequencing in the entire sample of individuals is preferable as it allows for an unbiased determination of the contribution of rare variants to disease risk.

Despite the extensive *in vitro* work done as a part of this dissertation, further work can be done to characterize these variants *in vitro*. Additionally, it has yet to be
determined if the variants contributing to the observed associations have an effect on function and behavior \textit{in vivo}. In order to further test the functional effect of the observed variants in \textit{CHRNB4}, or other genes, in a more suitable context, it is now possible to derive induced pluripotent stem cells (IPSCs) from fibroblasts obtained from a skin biopsy. These IPSCs can then be differentiated into neurons by incubating the IPSCs with specific neuronal differentiation factors. These differentiated cells are capable of producing action potentials and express markers of various neuronal populations. IPSCs can be derived from individuals with and without a given genetic variant and the electrophysiological properties of their IPSC derived neurons can be determined. Further, current technology including zinc-finger nucleases, TALEN effectors and CRISPR-Cas9, can be used to alter a specific nucleotide base and compare the properties of cells with or without a mutation while retaining identical genetic background. This would likely reduce the signal to noise ratio as variance induced by comparing individuals with differing genetic backgrounds would no longer be an issue. In order to investigate the effects of observed variants \textit{in vivo}, a reasonable experiment would be to construct either lentivirus or adenovirus constructs that are able to infect live cells and express wild-type or variant forms of nicotinic receptor subunits as well as GFP as a marker of infection. Upon infecting interesting brain regions, for instance the medial-habenula or interpeduncular nucleus, slices could be made and electrophysiological measurements taken to determine the effect of variants on receptor function \textit{in vivo}. Additionally, living mice infected with these constructs can be subjected to nicotine conditioned place preference testing to determine if expression of variant receptors lead to altered nicotine preference or aversion. Knockout mice for the receptor subunit under investigation, in
this case \textit{CHRNB4}, would be preferable to wild-type mice as all observed CHRNB4
would necessarily be introduced via viral infection. A far more expensive alternative to
this lentiviral/adenoviral approach would be to create knock-in mice with only a single
amino acid changed and then perform the above-mentioned experiments on the resultant
mice. This can now be more readily done with the newly described TALEN or CRISPR-
Cas9 systems of genome editing, but remains far more expensive than the
lentiviral/adenoviral approach. This approach would be preferred generally over
expression of variant forms of \textit{CHRNB4} in specific brain regions as it does not suppose \textit{a priori} which brain regions are responsible for potential behavioral effects. Additionally,
resultant mice can be propagated without the need to grow viruses or perform stereotaxic
injections. One further disadvantage to this technique is that it is not as easy to investigate
a large number of variants in a short amount of time, a possibility when using the viral
expression technique.

There is much work to be done in the identification of new variants and genes
underlying risk for nicotine, alcohol and cocaine dependence and the characterization of
associated variants and genes in cellular and animal models. Both of these avenues of
investigation undertaken as part of this dissertation remain to be fully explored. It is my
hope that others in our laboratory and other groups will replicate the findings presented
here, expand upon them and utilize the insights drawn regarding the integration of
functional information into association analyses to identify other novel genes involved in
complex diseases.
References

36. Saccone, S. F., Hinrichs, A. L. & Saccone, N. L. Cholinergic nicotinic receptor


52. Xie, P. *et al.* Rare Nonsynonymous Variants in Alpha-4 Nicotinic Acetylcholine Receptor Gene Protect Against Nicotine Dependence. *BIOL PSYCHIATRY* 70,
71. Yang, B. Z., Han, S., Kranzler, H. R. & Farrer, L. A. A genomewide linkage scan of cocaine dependence and major depressive episode in two populations. ... (2011).
73. Lohoff, F. W., Bloch, P. J., Hodge, R., Nall, A. H. & Ferraro, T. N. Association analysis between polymorphisms in the dopamine D2 receptor (DRD2) and dopamine transporter (DAT1) genes with cocaine ... *Neuroscience*(2010).
74. Cao, J., LaRocque, E. & Li, D. Associations of the 5-hydroxytryptamine (serotonin) Receptor 1B gene (HTR1B) with alcohol, cocaine, and heroin abuse. *Journal of Medical Genetics Part B:* ... (2013).
76. Kedmi, M. & Beaudet, A. L. Mice lacking neuronal nicotinic acetylcholine receptor β4-subunit and mice lacking both α5-and β4-subunits are highly resistant to nicotine-induced seizures. *Physiological genomics*(2004).
86. Li, P., McCollum, M., Bracamontes, J. & Steinbach, J. H. Functional characterization of the α5 (Asn398) variant associated with risk for nicotine...
dependence in the $\alpha_3\beta_4\alpha_5$ nicotinic receptor. *Molecular ...* (2011).
87. Tammimäki, A. *et al.* Impact of human D398N single nucleotide polymorphism on intracellular calcium response mediated by $\alpha_3\beta_4\alpha_5$ nicotinic acetylcholine receptors. ... (2012).
96. Rezvani, A. H., Slade, S., Wells, C., Petro, A. & Lumeng, L. Effects of sazetidine-A, a selective $\alpha_4\beta_2$ nicotinic acetylcholine receptor desensitizing agent on alcohol and nicotine self-administration in selectively bred alcohol- ..... ... (2010).
97. Löf, E., Olausson, P., Stomberg, R. & McIntosh, J. M. Nicotinic acetylcholine receptors in the ventral tegmental area mediate the dopamine activating and reinforcing properties of ethanol cues. ... (2007).
101. Steensland, P., Simms, J. A. & Holgate, J. Varenicline, an $\alpha_4\beta_2$ nicotinic acetylcholine receptor partial agonist, selectively decreases ethanol consumption and seeking. in (2007).
102. Robles, N. & Sabriá, J. Effects of moderate chronic ethanol consumption on


105. Zarrindast, M. R., Meshkani, J., Rezayof, A. & Beigzadeh, R. Nicotinic acetylcholine receptors of the dorsal hippocampus and the basolateral amygdala are involved in ethanol-induced conditioned place preference. ... (2010).


113. Taslim, N., Soderstrom, K. & Dar, M. S. Role of mouse cerebellar nicotinic acetylcholine receptor (nAChR) α (4) β (2)-and α (7) subtypes in the behavioral cross-tolerance between nicotine and ethanol- …. *Behavioural brain research* (2011).


117. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix …. *Nucleic Acids*
177


134. Yoo, J. H., Chen, Y., Kitchen, I. & Bailey, A. The stereotypy-inducing and
OCD-like effects of chronic ‘binge’ cocaine are modulated by distinct subtypes of nicotinic acetylcholine receptors. ... of pharmacology (2012).


140. Rice, J. P. et al. CHRNA3 is more strongly associated with Fagerström test for cigarette dependence-based nicotine dependence than cigarettes per day: phenotype definition changes genome-wide association studies results. Addiction 107, 2019–2028 (2012).


