Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

_____ _____

Anna K. Wiste Date

Neuronal nAchR genes in smoking behavior and addiction

By

Anna K. Wiste
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Neuroscience Program

_____	_____
Clinton D. Kilts, Ph.D. Advisor	Thorgeir E. Thorgeirsson, Ph.D. Advisor
_____	_____
Elisabeth B. Binder, M.D., Ph.D Committee Member	Andrew P. Escayg, Ph.D. Committee Member
_____	_____
Allan I. Levey, M.D., Ph.D. Committee Member	Kerry J. Ressler, M.D., Ph.D.. Committee Member

Accepted:

_____
Dean of Graduate School Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Neuronal nAchR genes in smoking behavior and addiction

By

Anna K. Wiste
B.A. Columbia University, 2001

Advisors: Thorgeir E. Thorgeirsson, Ph.D.
Clinton D. Kilts, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the Graduate School of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in
Graduate Division of Biological and Biomedical Sciences
Neuroscience Program
2009

Abstract

Neuronal nAchR genes in smoking behavior and addiction
By Anna K. Wiste

The nicotinic acetylcholine receptor subunit gene cluster on chromosome 15 containing CHRNA5, CHRNA3 and CHRNB4 has been highlighted in several association studies of genetic influences on smoking and smoking-related diseases. We reported genome-wide significant association of the synonymous SNP, rs1051730, with smoking quantity, and also demonstrated that the T allele of this variant carried an increased risk of nicotine dependence, lung cancer and peripheral arterial disease.

Given the high rate of smoking among individuals with other substance dependence disorders, we sought to determine whether this variant also increased risk for dependence on other substances. Subjects met DSM criteria for dependence for at least one substance of abuse, including alcohol, amphetamine, cannabis, opiates, and sedatives/anxiolytics. Smoking quantity information was available from questionnaire data for a subset of subjects. Among alcohol dependent subjects the variant had the same effect on smoking quantity as observed in the general population.

To characterize the region, the role of variants within the region in regulation of gene expression was explored. Blood and adipose tissue from a large sample of subjects was used to test for association of the variant to expression of CHRNA5. Several correlated variants within the region show significant association to expression of CHRNA5.

The association with smoking and smoking-related disease reported in our study and others has been observed in genome-wide association studies. In most cases, association signals detected in genome-wide association studies are expected to be indirect, occurring through linkage disequilibrium with a functional variant. To identify the most likely functional variant responsible for the association signal we have performed sequencing in the gene cluster in 3 samples: nicotine dependence, lung cancer, and low quantity smokers. Among common variants in the cluster no variants were identified with stronger association than rs1051730. A functional SNP in CHRNA5, rs16969968, is in perfect LD with rs1051730 in Iceland, and it is the only common functional SNP in the cluster. Based upon these factors, rs16969968 is currently the variant most likely to be responsible for the association signals observed in this cluster to smoking behavior and smoking-related diseases.

Neuronal nAchR genes in smoking behavior and addiction

By

Anna K. Wiste
B.A. Columbia University, 2001

Advisors: Thorgeir E. Thorgeirsson, Ph.D.
Clinton D. Kilts, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements of the degree of
Doctor of Philosophy in
Graduate Division of Biological and Biomedical Sciences
Neuroscience Program
2009

Acknowledgements

The path I followed to reach my doctorate was not the usual path, and thus I must first thank and acknowledge the leadership of the MD/PhD program and the Neuroscience program, for their unflinching support in helping me to follow the best path for me, even if it was an unusual one. I must thank Allan Levey and David Rye for helping me to take that first leap of traveling to Iceland, and Thorgeir Thorgeirsson for being willing to take the chance of welcoming a graduate student. My Emory committee, starting with my advisor, Clint Kilts, and Allan Levey, Kerry Ressler, Elisabeth Binder, and Andrew Escayg similarly agreed to take on an unknown task, and they have all gone beyond the call of duty to support me in this endeavor, scheduling meetings on the spur of the moment, in unusual places, at difficult times. They did whatever it took so that this transcontinental experiment could work for me.

There are many, many people at Decode Genetics who should be acknowledged. First is Kari Stefansson who similarly took an unusual step in allowing an American graduate student to perform thesis work within the company. I hope I have proven the decision was a good one. Others I would like to acknowledge for their help are Thorunn Rafner, Thorlakur Jonsson, Augustine Kong, Daniel Gudbjartsson, Hreinn Stefansson, Braji Walters, Jelena Kostic, Theodora Thorlacius, Iris Honsdottir, Auslaug Jonsdottir, Adalbjorg Jonsdottir, Gisli Masson, amongst many others who provided time, materials, and technical work which was absolutely essential to the completion of this work.

There are three people I would like to mention individually. Andres Ingason started to help me on the first day I began working at Decode, and his help in gaining my feet was essential. Later on he was a valuable sounding board for ideas and theories. If we could not prove each other's theories wrong, then perhaps we had actually hit on something. Later in my work Gudrun Anna Jonsson was an important support for me as we worked closely on the addiction projects at Decode. Having recently completed her PhD she could commiserate, but more importantly, she was also a sounding board, able to rein me in from some of my ideas which, while exciting, could tend towards unrealistic. She has been a very good friend to me. Frank Gellar is the reason why any statistics provided in this thesis are correct, and any faults are my own. Frank took time to explain to me in detail the tests we used and why, no matter how many times it took for the perplexed look to leave my face.

I would like to thank Clint Kilts, my co-advisor, who has known me well for many years and served as mentor in science and in life before, through, and after my thesis work.

Lastly and most importantly, is Thorgeir Thorgeirsson, my co-advisor. Thorgeir took a chance in blindly accepting to take on for even three months a graduate student from the US who had not yet found her scientific home. It turned out to be a perfect match. I could not have asked for a better advisor. He is brilliant and innovative, and capable of explaining everything from the most basic concepts of human genetics to the most complicated new methods we tried to develop with equal aplomb. Without him this thesis certainly would not exist. The value of the knowledge I gained from him about science is surpassed only by the knowledge I gained from him about life.

Table of Contents

Figures and Tables

Chapter 1

General Introduction

Addiction is a common and costly public health problem. The most commonly used substance of abuse is nicotine. Tobacco use is a leading cause of preventable death, contributing to cancer, cardiovascular and pulmonary disease, accounting for over 400,000 deaths in the US each year(1), and over 5 million worldwide(2). The prevalence of nicotine dependence in the US is 12.8%(3). In 2006, the smoking rate in the US was 20.8%. Of the current smokers, 44% had attempted to quit smoking at least once in the previous year, underscoring the difficulty of successful smoking cessation(4). The identification of a genetic factor which significantly influences smoking opens a door which may lead to a better understanding of the effects of nicotine in the brain and body, perhaps paving the way for more effective treatment strategies. Individuals with other substance abuse problems are a significant fraction of nicotine dependent smokers. A better understanding of the role of smoking in these disorders is important for the successful treatment of nicotine dependence and dependence to other substances of abuse in these individuals.

A. Genome-wide Association Studies and Their Impact on Genetics of Complex Disease

Within the last two years genome-wide association studies have had a tremendous impact on our understanding of the genetics of complex disorders. The study of human genetics in Iceland has provided an unusual opportunity to take advantage of this new technology, and at Decode there has been great success using this method(5-11) in a wide variety of both diseases such as atrial fibrillation(11) and prostate cancer(9), and traits such as pigmentation(7). Large sample sizes are available for common diseases in affected groups. Moreover, the availability of DNA samples from a large proportion of

the population allows for very accurate estimations of the population frequency of variants tested in these studies. Genetic homogeneity minimizes population stratification effects and extensive genealogical information allows the use of strategies which rely on relatedness among affecteds, such as linkage analysis, while also providing the opportunity to control for relatedness and make proper corrections in association analysis where it is a confound.

Many other studies have also made successful use of this technique. A large case-control study looking at multiple unrelated diseases identified risk variants for Crohn's Disease, Type 1 diabetes, and additional variants for Type 2 diabetes(12). Other studies identified risk variants for colorectal cancer(13, 14), celiac disease(15) and systemic lupus erythematosis(16) among others. Multiple studies have also looked at genetic variant influencing height, finding a surprisingly large number(17-19). These studies have also presented new challenges for data analysis and interpretation. The key role replication studies play in gene discovery, and the need for publication of both positive and negative results, has been stressed(20). Multiple testing creates the need for application of rigorous standards in determining statistical significance. A Bonferroni correction, for the number of markers tested, remains the best method, although it is conservative. Both population stratification and relatedness among study subjects also need to be taken into account. The method of genomic controls is employed to correct for these elements(21). As initial genome-wide association studies have been completed, meta-analyses are also beginning, with non-insulin dependent diabetes as an excellent example. In a meta-analysis of three scans including over 10,000 individuals, with a large number of additional cases for replication, 6 additional variants were identified which contribute significant risk for non-insulin dependent diabetes(22). This brings the total number of variants identified up to

about 20(23). This example highlights the complexity of the genetics of common disorders, and how genome-wide association studies can be used in multiple ways to increase their ability to detect variants which confer small relative risks of disease. Meta-analysis or joint analysis of independent data sets in addition to replication studies for individual findings have demonstrated success at identifying significant risk factors not identified by individual studies(22).

Prior to genome-wide association studies the common approach to study of complex disorders was positional cloning. This approach did yield some successes(24, 25), but was far more successful in the identification of genes responsible for Mendelian disorders(26, 27).

Several important developments in the understanding of the human genome and the principles underlying its variation have been key to allowing these advances. The first step in this process was the sequencing of the human genome, the first draft of which was released in the year 2001(28). A high resolution genetic map of microsatellites has improved precision in linkage analysis, and was helpful in aiding the assembly of the genomic sequence(29). While the Human Genome Project obtained the full sequence of the human genome, the HapMap project directed attention to characterizing the variation in the human genome. It has built on the general consensus sequence to identify and characterize the common variations throughout the genome(30). Genotype data on these variants can be used to determine linkage disequilibrium (LD), a measure of correlation, between SNPs. Two measures of LD are most used and both have important utility. The first one, D', is most useful for determining large regions, or blocks, in which all variants are closely related and there have been few recombination events within the region(31). As long as one allele of a SNP only appears with a particular allele of a second SNP,

D'=1. The second measure used is r^2, which is used to determine the ability of SNPs to effectively tag other surrounding SNPs(32). If two SNPs are exactly equivalent then r^2=1.

Association studies all rely on this concept. Rather than testing all possible variants in a region, or in the case of genome-wide association all variations within the entire genome, a subset of SNPs can be genotyped that provides representative information on other variants in the region. Once a SNP has been identified through this method the association may be indirect and careful testing of other variants in the region is necessary to identify the actual risk-conferring variant. The HapMap project has been essential to defining the LD structure of the human genome, and has highlighted differences in this structure between populations of different ethnic backgrounds. When the LD structure is different between two populations an association signal in one population may be correlated with a functional risk variant in that population but may not be correlated in the second population. This could lead to non-replication despite the presence of a true risk variant. If however, an association signal does replicate in a population with a significant difference in LD structure within the region, powerful evidence is present for a true association. The available genotyping chips tag approximately 85% of comon SNPs in Caucasian populations, while rare variants may not be tagged at all(33).

B. Definition of Phenotypes

While several measures can be used to examine smoking behavior and its genetic basis, two systems of criteria are most commonly used to define nicotine dependence in research. In the Diagnostic and Statistical Manual (DSM-IV)(34) the 7 criteria for substance dependence are adapted and used for nicotine dependence (See Figure 1.1). To

meet DSM-IV criteria the subject must experience 3 or more of these criteria within the same 12 month period. The Fagerstrom Test for Nicotine Dependence (FTND) was developed specifically to measure nicotine dependence through addressing smoking behavior. The revised FTND(35) consists of 6 questions, leading to a score ranging from 0 to 10 (See Figure 1.2). Different studies have used different cut-off points on this scale to define affected status for nicotine dependence. Previous genetic studies of nicotine dependence have used 4 or higher as the definition(36-38), and we have followed suit. However, the choice of an appropriate cutoff point is not clear. One linkage scan using the earlier Fagerstrom Tolerance Questionnaire (FTQ), from which the FTND was derived, used a cutoff of 7, which would be approximately equal to a 6 on the FTND(39).

The FTQ and FTND scales were designed to be used as a continuous variable, and to some extent, as a contrast to dichotomous systems like the ICD and DSM(40). FTND scores compiled from studies in population samples from 13 countries yielded mean scores of 2.8-4.6 in regular smokers(41). In this study, mean FTND score was inversely correlated with a country's smoking prevalence, suggesting that as prevalence reduces, the remaining smokers are those with higher nicotine dependence. The developers of the scale do not themselves use it to create a dichotomous variable(35, 40, 41). One paper discussed in detail later in this chapter uses 6-10 as high nicotine dependence and 0-4 as low nicotine dependence(42). Thus, individuals with a score of 4 would have appeared in their control group. A higher cutoff point on the scale undoubtedly ensures a more severely affected case group. However, there is also evidence that higher FTND scores correlate with substance use problems and other psychopathology(43), an effect also seen in analysis of Decode questionnaire data. Limiting a nicotine dependence group to those most severely affected may enrich the sample for subjects with other psychiatric disorders

or symptomatology in such a way that the sample is no longer representative of the general population of smokers. A balance must be struck, and a consensus does not seem to have been reached.

Phenomenologically, there is little overlap between the components of addiction examined by these two systems. The FTND scale is heavily influenced by items addressing cigarettes per day smoked and the time to first cigarette after waking. Meanwhile the DSM criteria have been developed for alcohol and illicit substance use and the DSM system has a broader approach including psychological aspects of addiction. Thus each system has a different emphasis. The FTND is specific to the unique aspects of nicotine use, but does not address psychological components of addiction. The DSM criteria apply a broad approach to addiction as a brain disorder with several different elements, but do not take into account the distinctive aspects of nicotine dependence in comparison to other substance dependencies.

The Decode Nicotine Dependence study uses a questionnaire on smoking behavior which fully addresses all of the criteria for both systems, and has been translated into Icelandic. The questionnaire has been administered to over 5000 subjects.

The most commonly used definition of substance dependence in research studies is the endorsement of 3 or more DSM criteria. The majority of subjects in our studies with other substance dependence diagnoses have been treated for substance use disorders at SAA, the largest addiction treatment center in Iceland. They have been diagnosed at admission according to DSM criteria by clinical staff.

Additional subjects are included in our analyses who participated in a family-based study of anxiety disorders and depression. The Composite International Diagnostic Interview (CIDI)(44) was used in this project with the full substance use module included

for all individuals. The CIDI is designed to allow diagnosis according to ICD-10 criteria, but DSM-based diagnoses can also be extracted. Subjects meeting criteria for any substance dependence based on these clinical interviews are also included.

Other smoking phenotypes are also considered in genetics research aside from the concept of nicotine dependence. Smoking behavior is sometimes divided into smoking initiation and smoking persistence. Conceptually these two key elements of smoking behavior are easily understood. In practice they can be difficult to define. Smoking initiation refers to the process of transition from non-smoker to smoker. It is generally defined in some way that involves progression to regular smoking(45). In our case, we use the question 'Have you ever smoked regularly for at least 1 year?' Other phrasings would include 'Do you smoke or have you at some time smoked regularly, in other words daily or almost daily?' and 'Have you EVER been a smoker?'(45). The second element of smoking behavior is persistence, or continued smoking. It can be defined simply as being a current smoker at the time of interview or in terms of length of time smoked. Smoking quantity or quantity of cigarettes smoked per day is another variable which can be used to study smoking behavior. In our study we ask 'How many cigarettes per day do you smoke or did you smoke?' Maximum cigarettes smoked in one day is yet another smoking behavior variable used(46).

There is a well known connection between alcoholism and smoking (47). The smoking rate in alcohol dependence is estimated at 80%(48), in line with the 87% smoking rate at admission among patients at the largest treatment center in Iceland (SAA)(49). The National Epidemiologic Survey on Alcohol and Related Conditions found rates of alcohol and drug dependence 4 fold higher in subjects with nicotine dependence, and an approximately 4 and 6 fold increase in prevalence of nicotine

dependence among subjects with alcohol and drug dependence diagnoses, respectively(3). The reasons for this connection are not clear, but there are several possibilities. While social factors are likely to play a role in the concurrent use of the two substances, there is also evidence of biological factors. Use of both drugs influences similar mechanisms within the dopaminergic reward system which may lead to abuse. There is also evidence that a person's subjective responses to alcohol are influenced by consumption of nicotine and vice versa, and last, there are genetic studies which suggest shared liability for abuse of both substances(50). Individuals with alcohol dependence and nicotine dependence may require different approaches to smoking cessation in comparison to the general population. There is a need for better understanding of these interactions in order to provide such targeted assistance(51).

Figure 1.1

DSM-IV Criteria for Substance Dependence

A maladaptive pattern of substance use leading to clinically significant impairment or distress, as manifested by three (or more) of the following, occurring any time in the same 12-month period:
1. Tolerance, as defined by either of the following:
(a) A need for markedly increased amounts of the substance to achieve intoxication or the desired effect
or
(b) Markedly diminished effect with continued use of the same amount of the substance.

2. Withdrawal, as manifested by either of the following:
(a) The characteristic withdrawal syndrome for the substance
or
(b) The same (or closely related) substance is taken to relieve or avoid withdrawal symptoms.

3. The substance is often taken in larger amounts or over a longer period than intended.

4. There is a persistent desire or unsuccessful efforts to cut down or control substance use.

5. A great deal of time is spent in activities necessary to obtain the substance, use the substance, or recover
from its effects.

6. Important social, occupational, or recreational activities are given up or reduced because of substance use.

7. The substance use is continued despite knowledge of having a persistent physical or psychological problem that is likely to have been caused or exacerbated by the substance (for example, current cocaine use despite recognition of cocaine-induced depression or continued drinking despite recognition that an ulcer was made worse by alcohol consumption).

Nicotine Withdrawal Syndrome

A. Daily use of nicotine for at least several weeks.
B. Abrupt cessation of nicotine use, or reduction in the amount of nicotine used, followed within 24 hours by four (or more) of the following signs:
    1. dysphoric or depressed mood
    2. insomnia
    3. irritability, frustration, or anger
    4. anxiety
    5. difficulty concentrating
    6. restlessness
    7. decreased heart rate
    8. increased appetite or weight gain
C. The symptoms in Criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Taken from DSM-IV TR(34)

Figure 1.2

Fagerström Test of Nicotine Dependence (FTND) **(Points)**

1- How soon after you wake up do you smoke your first cigarette?

Within 5 minutes	(3)
6-30 minutes	(2)
31-60 minutes	(1)
after 60 minutes	(0)

2- Do you find it difficult to refrain from smoking in places where it is forbidden, e.g. at the mosque (church), at the bus?

Yes	(1)
No	(0)

3- Which cigarette would hate most to give up?

The first one in the morning	(1)
All others	(0)

4- How many cigarettes/day do you smoke?

10 or less	(3)
11-20	(2)
21-30	(1)
31 or more	(0)

5- Do you smoke more frequently during the first hours after waking than the rest of the day?

Yes	(1)
No	(0)

6- Do you smoke when you are so ill that you are in bed most of the day?

Yes	(1)
No	(0)

Taken from Heatherton et al.(35)

C. Genetic Basis of Smoking Behavior and Nicotine Dependence

In order to be able to identify genes which confer risk for nicotine addiction, a significant genetic component to risk for the disease must first be proven. Li et al. (52) performed a meta-analysis of twin studies including both the phenotypes of smoking intiation and persistence. Heritability estimates were different for men and women. For smoking initiation h^2 was .37 for males and .55 for females. For smoking persistence, h^2 was .59 for males and .46 for females. In a Dutch population, Vink et al(53) examined heritability for both initiation and nicotine dependence. Heritability estimates did not differ between males and females. For smoking initiation, h^2 equaled .44, while in nicotine dependence, h^2 equaled .75. The particularly high heritability estimate for nicotine dependence, as defined by the Fagerström Test for FTND suggests that the FTND scale may be a more useful measure for identifying genetic elements influencing smoking than other measures of smoking behavior.

Recent studies have also looked at genetic components of nicotine dependence and smoking in the context of other addictions. Much of the genetic risk of addiction appears to be shared across substances, but amongst the most common substances of abuse, nicotine is unique in appearing to have a substantial proportion of its genetic risk for dependence accounted for by substance-specific effects rather than general factors contributing to risk for dependence to all substances of abuse(54). There are also discernable genetic effects distinct to the more general phenotypes of smoking quantity and alcohol consumption individually, but also evidence of genetic factors which affect both(55).

D. Nicotinic acetylcholine receptors – structure and role in nicotine dependence

Nicotinic acetylcholine receptors are made up of five subunits, which combine to form a pore that becomes a non-specific cation channel. Neuronal nicotinic acetylcholine receptors (nAchR) are made up of combinations of the subunits α2-10 and β2-4. Higher affinity receptors are made up of α2-6 subunits in combination with β2-4 subunits. There are 2 binding sites for acetylcholine in nAChRs, each at the junction of an α and a β subunit. Lower affinity receptors are homomeric pentamers of α7-10 subunits. These receptors have 5 binding sites, one at each subunit. The properties of the receptors are influenced by the subunit combination, including affinity, ion selectivity, and desensitization. Two subunits are purely modulatory, the α5 and β4 subunits. These only occur in combination with other subunits from their group(56).

Given the importance of dopaminergic reward pathways in addiction, the cholinergic innervation of the ventral tegmental area (VTA) has been well studied(57, 58). The nAchRs located on dopaminergic neurons themselves are primarily of the α4/β2 combination (59). Similar receptors are also located on GABAergic neurons in the VTA(60). Glutamatergic terminals are also present in the region bringing feedback from cortical areas, and the primary nAchR subtype on these terminals is the α7(61). Evidence indicates that at concentrations of nicotine reached in the brain after smoking cigarettes, the receptors located on dopaminergic and GABAergic terminals desensitize rapidly(60), while the lower affinity receptors on glutamatergic terminals do not. This combination leads to a decrease in GABAergic tone and an increase in glutamatergic input as leading to an increase in dopamine release from VTA neurons. The release of dopamine in the nucleus accumbens from VTA neurons is a shared effect on the brain of substances of

abuse(62) The release of dopamine from the VTA is key for the rewarding properties of these substances which can then contribute to development of dependence(63).

E. Identification of Risk Variant in *CHRNA5/CHRNA3/CHRNB4* gene cluster

The *CHRNA5/CHRNA3/CHRNB4* gene cluster was first highlighted in a candidate gene approach using genome-wide association data(38), from the first genome-wide association study of nicotine dependence(37). This study used a two-stage design in which 2.4 million SNPs were initially tested in pooled DNA samples. In the second stage, 31,960 SNPs were selected and genotyped on individual samples from the original discovery cohort and additional cases and controls. 348 genes were chosen as candidate genes, and 3713 SNPs from these genes were analyzed separately giving weight to the SNPs based on the strength of the evidence leading the authors to designate the gene as a candidate gene(38). The non-synonymous SNP, rs16969968 in *CHRNA5* was highlighted in the candidate gene portion of the study as having the most biologically compelling evidence of a genetic risk factor, based on its potential functional impact and its position among the top association signals. In the same study rs1051730 was also genotyped, and, in strong LD with rs16969968 gave a similar signal. No association signals reported in this study were statistically significant on their own. The study used 1050 cases (defined as FTND≥4) and 879 controls which were chosen as having smoked at least 100 cigarettes but having a lifetime FTND score of 0. The selection of controls tightened the focus of the study. All subjects had exposure to the substance, and the comparison was between those who developed dependence and those who did not.

The next report to bring attention to this region was a genome-wide association study of smoking quantity, using a QTL approach(64). Two large population samples of

smokers were genotyped with SNP arrays. In both samples SNPs located in the *CHRNA5*/*CHRNA3*/*CHRNB4* cluster were the strongest signals. Neither rs1051730 nor rs16969968 were genotyped in either sample included in this study. However, the SNPs exhibiting association were all in LD with rs16969968 and rs1051730, and most likely were detecting the same signal. No marker showed statistically significant association after correction for multiple testing.

Three studies were published on the same day which put the importance of the cluster beyond any doubt. Two large genome-wide association scans for lung cancer reported that the only variants to achieve genome-wide significance for lung cancer were variants within this block, including rs1051730 and additional markers in close linkage disequilibrium with it(65, 66). Both studies matched control groups for smoking behavior in order to control for the confounding variable of high lifetime smoking exposure.

In the third study, we reported our genome-wide association study of smoking quantity, using a QTL approach in which smoking quantity was divided into 4 categories (0-10, 11-20, 21-30, and 31+)(67). This analysis included 15,771 subjects with smoking quantity information gathered through participation in various projects at Decode genetics and replication samples from Spain and The Netherlands. Association of smoking quantity with rs1051730 was highly significant ($P=6\times10^{-20}$, N=13945) We also found significant association with nicotine dependence, defined as either meeting DSM criteria or scoring 4 or more on the FTND scale (OR=1.17, 95%CI:1.10-1.25, $P=3.3\times10^{-6}$). This association was much stronger (OR =1.40, $P=7\times10^{-15}$) when a control group comparable to that used in the Saccone et al.(38) study was used for comparison. While full FTND scores were not available for the control group, we used subjects who reported regular

smoking (including social smoking) for at least a year but never reported smoking more than 10 cigarettes per day.

We further observed significant association of the variant to lung cancer and peripheral arterial disease in Iceland and several foreign cohorts (LC: OR=1.31, $P=1.5\times10^{-8}$; PAD: OR=1.19, $P=1.4\times10^{-7}$). Lung cancer and peripheral arterial disease are two diseases with very strong relationships to smoking. Smoking is the major risk factor for lung cancer (LC)(68-71), and one of the main risk factors for peripheral arterial disease (PAD)(72-74). Thus the finding of significant association with both of these disorders is at least partly due to the variant's effects on smoking behavior. However, the risk for lung cancer attributed to this variant is higher than would be expected based on the effect of the variant on smoking behavior as measured. The measures used for smoking behavior here cannot be expected to fully capture the true effect of the variant on smoking. A reponse to a questionnaire item on cigarettes smoked per day at a single timepoint is an inherently noisy measure. The dataset however is so large that significant effects can be determined with this measure. Thus, the difference between the observed risk of lung cancer and that expected based on the smoking quantity effect can have multiple explanations. The difference can be due to a feature of smoking behavior which this variant increases risk for, and which is important in risk for lung cancer, but is not captured by the smoking quantity question. A second possibility is that the variant itself confers additional specific risk for lung cancer in addition to its effect on smoking behavior. A third possibility is that another variant, in linkage disequilibrium with rs1051730, has a specific effect on lung cancer which is captured by rs1051730 and appears as added risk through linkage disequilibrium. Further detailed genetic and

functional studies are needed to determine with confidence which of these possibilities is actually at work.

F. Additional findings on the CHRNA5/CHRNA3/CHRNB4 gene cluster variant

Since the identification of the risk variant, several replications have been reported in independent populations(42, 75, 76). Bierut et al.(75) used a family-based association test to analyze the variant in European American families from the Collaboraotive Studies on the Genetics of Alcoholism (COGA) project and found significant association with habitual smoking. The study also suggests that there is an independent risk factor in the same linkage disequilibrium block, rs578776. Weiss et al.(42) found significant association of rs1051730 with nicotine dependence according to the FTND scale. They used a score of 0-4 as controls and 6-10 as affecteds. However, in their sample, association was only significant for smokers who began smoking before the age of 17. Spitz et al.(76) have followed up on their genome-wide association study of lung cancer(66) by studying the association of the variant with smoking in their lung cancer case-control cohort. The T allele was associated with higher FTND scores and increased smoking quantity in both lung cancer cases and controls.

The Spitz et al study(76) also provided further data on the role of the variant in lung cancer. A large series of never-smoking lung cancer cases and controls were tested and no association was found with the variant, indicating that the risk of this variant for lung cancer among smokers is contributed through nicotine, whether directly or indirectly.

In addition, two studies have examined the role of the cluster in other substance dependencies: alcohol(77), and cocaine(78). No association with alcohol dependence was seen for rs1051730, but nominal significance was seen with other variants in the block. In

a study of cocaine dependence, association was observed for rs1051730 and rs16969968, however it was in the opposite direction of that seen for nicotine dependence, as the T allele of rs1051730 appeared to be protective (OR=0.67, P=0.005).

There is still much work to be done to characterize the role of the variant in smoking and its interactions with other phenotypes, including other addictions, psychiatric disorders and diseases for which smoking is a risk factor, such as pulmonary and cardiovascular disease.

G. Previous Characterizations of the Region

Three studies have reported on characterization of the genes in this cluster. Initial characterizations of the cluster focused on the genes as candidates for rare mendelian disorders, including megacystitis-microcolon-hypoperistalsis syndrome (*CHRNA3*/*CHRNB4*)(79) and autosomal dominant nocturnal frontal lobe epilepsy (*CHRNA5*/*CHRNA3*/*CHRNB4*)(80, 81). In each case no mutations identified were found to be causative for the diseases studied. However, valuable early information on the structure of the three genes and the common genetic variants within them was provided by these studies.

The Weiss et al study(42) which replicated the findings of association to nicotine dependence also involved sequencing of the region. Three groups were chosen for sequencing, including 72 subjects with high nicotine dependence, 72 with low, and 48 non-smokers. The study covered all nicotinic acetylcholine receptor subunits, including the three in the cluster on chromosome 15q25. The aim of the sequencing in this instance was to identify tagging SNPs for association testing in a larger sample. Once tagging

SNPs were taken forward into the larger group, only the variant rs1051730 and highly correlated SNPs carried significant association with nicotine dependence.

H. Functional Importance of Genes in the Cluster

The nicotinic acetylcholine receptor (nAchR) subunits encoded by genes within this cluster have not been the focus of the bulk of existing research on nAchRs in addiction. The α4, β2, and α7 subunits have received far more attention. However there are some studies available that give insight into the role these receptors play in the body, and in response to nicotine in particular.

The α3 subunit has a limited expression profile in the brain and indeed in the body. A comprehensive examination of expression in rat brain demonstrated strongest expression throughout the thalamus. Other areas with significant expression include portions of entorhinal cortex and the locus coeruleus, as well as cranial nerves. Some expression was also identified in the ventral tegmental area(82). Outside the brain, areas with significant expression include small and large intestines and bladder. The knockout mouse suffers early death, but the exact cause is not known. About 40% die in the first 3 days, with the remainder dying within 2 months after weaning. The most demonstrable phenotype of this mouse has to do with bladder abnormalities(83). Physiological studies have further shown a lack of contractility in response to nicotine in these mice(84). Use of the knockout mouse to explore the role of α3 subunits role in neural transmission have indicated an important role in the sympathetic nervous system. Fast excitatory transmission was found to be abolished between pre-and post-ganglionic neurons in the sympathetic nervous system when α3 subunits are not present(85).

The α5 subunit also has limited regions of strong expression in the brain, though these regions include the substantia nigra and ventral tegmental area and the subiculum and parasuciculum areas of the hippocampal formation(86). The α5 knock-out mouse performs normally on standard behavioral tests and is generally healthy. They are less sensitive to the effects of nicotine as demonstrated by decreased susceptibility to nicotine induced seizures(87). These mice are particularly susceptible to experimental colitis, suffering a significantly more severe phenotype than wild-type mice when colitis is pharmacologically induced. However, nicotine attenuated the effects in both wildtype and α5 knockout mice, indicating that the helpful effects of nicotine in ulcerative colitis are mediated by nicotinc receptor subunits other than the α5 (88). The role of α5 subunits is particularly difficult to determine because it only occurs as a modulatory subunit. Research on α4/β2 receptors indicates that when an α5 subunit is included the calcium permeability of the channel is higher, and the receptor has higher affinity and lower threshold for desensitization(89).

The β4 subunit has a wider expression pattern throughout the brain than either α5 or α3 subunits. The strongest expression is located in the medial habenula, but moderate levels are found throughout areas of the hippocampus and cortex. Brainstem expression is more limited, but significant expression is found in the locus coeruleus, and expression is also found in the cerebellum(90). The β4 knockout mouse also survives to adulthood with no gross abnormalities as with the α5 knockout(83). Most relevant to the study of addiction, β4 knockout mice have attentuated somatic withdrawal symptoms when mecamylamine injection is given following chronic nicotine administration(87).

I. Objectives

The T allele of the variant rs1051730 has been identified as contributing significant risk for nicotine dependence and correlates with smoking quantity, and this finding is replicated. That genetic factors influence addiction in general has been demonstrated, and susbtance-specific effects have also been postulated. It is important to screen risk variants identified for one addiction in other addictions to determine whether they influence addiction in general, or only addiction to a particular substance.

Chapter 2 describes a screen of the linkage disequilibrium block containing rs1051730 for association to other substance dependence diagnoses using genotype data available from chip genotyping. Given high rates of smoking and nicotine dependence and the increased nicotine consumption seen in alcoholics, it is also of interest to examine the effect of a smoking variant in this high risk population compared to the general population.

Chapter 3 describes a characterization of the region which was done to look for other sequence variants in the region which might better explain the association observed. Chapter 3 also includes a study of genetic regulation of the expression of CHRNA5 *in vivo* using samples from both blood and adipose tissue.

Chapter 2

Association of Smoking Quantity with a CHRNA5/CHRNA3/CHRNB4 variant in Substance Dependence, but no Association of Variants within the Cluster to Substance Dependence itself.

Material from this chapter also appears in an article which will be submitted to Molecular Psychiatry. Contributing authors of this article include: Anna Wiste, Frank Geller, Gudrun A. Jonsdottir, Gyda Bjornsdottir, Valgeir Runnarsdottir, Hogni Oskarsson, Thorarinn Tyrfingsson, Thorgeir Thorgeirsson, Kari Stefansson

A. Abstract

Rates of smoking and cigarette consumption among patients with alcohol and drug dependence are markedly higher than in the general population. The SNP rs1051730, within the *CHRNA5/A3/B4* gene cluster on chromosome 15, associates with nicotine dependence and smoking quantity, as well as smoking-related diseases. A recent study of alcohol dependence indicated no association with this SNP, but suggested association with another SNP within the cluster. We have screened the gene cluster for association with alcohol and drug dependence. No association was seen between alcohol dependence (n=2215) and any of the SNPs studied when compared to population controls (n=31,576). There was also no association between any SNP and dependence on amphetamines, cannabis, opioids or sedatives, anxiolytics or hypnotics. Data on smoking quantity is available on a subset of these subjects. When we examined association between smoking quantity and the SNP rs1051730 within the alcohol dependence sample, a significant effect is present (p=0.002, n=1691), and the effect size (approximately 1 cigarette per day per copy of the T allele of rs1051730) is similar to that observed in population controls.

B. Introduction

The strong association between alcoholism and smoking has been well established(47). The smoking rate in alcohol dependence is estimated at 80%(48), in line with the 87% smoking rate at admission among patients at the largest treatment center in Iceland (SAA)(49). Results from the National Epidemiologic Survey on Alcohol and Related Conditions demonstrate the high comorbidity between nicotine and other substance dependence, with rates of alcohol and drug dependence approximately 4 fold higher in subjects with nicotine dependence, and an approximately 4 and 6 fold increase in prevalence of nicotine dependence among subjects with alcohol and drug dependence diagnoses, respectively(3). Family-based studies have clearly demonstrated genetic components for alcohol(91) and drug dependence(92), as well as for smoking(52) and nicotine dependence(53), separately. A recent study demonstrated a strong genetic effect on the correlation between nicotine and alcohol consumption in adulthood(55). Another study from the same population demonstrated shared genetic components to risk for dependence to different psychoactive substances, but also suggested a substantial substance-specific effect for nicotine dependence(54). These findings highlight the importance of examining the role of genetic risk factors for one substance dependence in other addictive disorders.

Association of variants within the *CHRNA5*/*CHRNA3*/*CHRNB4* gene cluster has recently been reported with smoking quantity(64, 67), nicotine dependence(38, 67), and the smoking-related diseases lung cancer(65-67) and peripheral arterial disease(67). Association with nicotine dependence has been replicated in an independent Caucasian population(75). One report has suggested that association with nicotine dependence is

only present for those who start smoking regularly at an early age(93). The cluster has also been examined for association with alcohol(77) and cocaine(78) dependence. No association was seen with the nicotine dependence risk variant in alcohol dependence, while the nicotine dependence risk variant was reported to be protective for cocaine dependence.

In this study we have performed an association analysis of the cluster in a sample from the largest addiction treatment center in Iceland. We report results for alcohol, amphetamine, cannabis, opioid, and sedative, anxiolyic or hypnotic (SAH) dependence. Cocaine use is not prevalent in Iceland, and our sample size for cocaine dependence is currently not large enough to include in this study. We have also analyzed the association of the variant with smoking quantity among alcohol dependent smokers.

C. Materials and Methods

Subjects

Subjects for this study have been recruited primarily through SAA, the largest addiction treatment center in Iceland. All patients were admitted for an inpatient rehabilitation program, and diagnosed by clinicians according to DSM criteria at admission. Alcohol dependence is by far the most common diagnosis (n=1874). Other diagnoses include amphetamine (n=282), opioid (n=196), cannabis (n=374), and SAH dependence (n=357). Most of these subjects also have an alcohol dependence diagnosis. Additional subjects were recruited through a family-based study of anxiety and related disorders(94). Through this project the CIDI(44) was administered to 3100 individuals, with 561 receiving a diagnosis of alcohol dependence either by ICD-10 or DSM-III-R criteria. Within this group, 381 subjects received the diagnosis under both sets of criteria. Of the 561 subjects

from this group 238 have also been admitted to SAA for treatment. Smoking quantity information is available for many subjects (n=1691) through completion of questionnaires reporting cigarettes per day smoked. If subjects were current smokers they were asked to report on current smoking habits, while if they had quit, they were asked to report on past smoking habits. The study protocols were approved by the National Bioethics Committee (NBC) and the Data Protection Authority (DPA) of Iceland. The DPA has encrypted all personal identifiers linked to phenotype information or blood samples using a third-party encryption system(95). All subjects are of Icelandic ancestry. The control population consists of all subjects genotyped with Illumina chips as part of ongoing studies at Decode Genetics. All subjects with known substance use disorder diagnosis or nicotine dependence were removed from the control group prior to analysis. Of the 12,932 individuals analyzed for smoking quantity here, 9821 were included in our earlier study of smoking quantity and rs1051730. The controls used in this study include 23,859 individuals used previously as controls for nicotine dependence, lung cancer or peripheral arterial disease(67). See Table 2.1 for demographic information.

Genotyping

Genotyping was performed using Human Hap300 and Human Hap300-duo1 Bead Arrays and specialized software (Illumina; San Diego, CA, USA)(96). Only subjects with overall genotype yield greater than 98% were included in analyses, and individual SNPs with genotype yield less than 99% or significant deviations from Hardy-Weinberg equilibrium were excluded from analysis.

Association Analysis

We focused our association analysis on the LD block containing rs1051730 and the SNPs with reported association to alcohol dependence(77). The block was defined as described previously (9) using the definition of recombination hotspots from the snpRecombHotspotHapmap track of the UCSC genome browser (hg16 build 34)(97, 98). A block is defined as the region between the boundaries of two consecutive recombination hotspots. The block under investigation extends over *CHRNA5* and *CHRNA3* and the 3' end of *CHRNB4*, as well as *PSMA4* and *IREB2*, and contains 13 SNPs on the Illumina chips. Data from an additional 16 SNPs upstream and downstream of this region were also analyzed and the results are in the Supplementary Material.

Regression Analysis

As in our previous study(67), regression models were used to examine the effect of rs1051730 on smoking quantity. Information on cigarettes per day smoked was categorized into SQ levels and used as a quantitative variable. Four categories were used: (1-10), (11-20), (21-30) and (>30). Sex and year of birth are significant predictors of smoking quantity and have been included in the model. Year of birth is included as a categorical variable, with subjects divided into 4 birth year cohorts: ≤1940, 1945 to 1955, 1960 to 1965, and ≥1970. These categories divide the individuals with dependence diagnoses into groups of approximately equal size.

Statistical Analysis

A likelihood ratio test was used for analysis using χ^2 statistics. A correction factor has been applied using genomic control(21) based on genotypes from all SNPs included on both chips used and passing quality control measures. The χ^2 statistic is divided by the

mean χ^2 statistic for all 309,690 SNPs genotyped. This procedure takes into account relatedness among our subjects due to the fact that some recruitment has been family-based. Correction factors for the substance dependence diagnoses were 1.228,1.088, 1.081, 1.027 and 1.064, for alcohol, amphetamine, cannabis, opioid, and SAH dependence, respectively. The p-values reported do not correct for multiple testing, either of number of SNPs or phenotypes tested.

D. Results

Association analysis of CHRNA5/CHRNA3/CHRNB4 gene cluster in alcohol and other drug dependence

There is no significant association between alcohol dependence (n=2215) and any of the SNPs tested, including rs1051730 (OR=0.98, 95% CI=0.91-1.05) (See Table 2.2). Our treatment center sample includes smaller subsets with amphetamine, opioid, cannabis, and SAH dependence diagnoses that in most cases have also been diagnosed with alcohol dependence. No significant association was detected between any of the SNPs tested and these other substance dependence diagnoses (See Table 2.3).

The SNP rs588765 has been suggested for further study in alcohol dependence(77), based on association results in the COGA sample and an African-American sample. In the latter sample the LD of rs588765 to rs6495306 is reported as D'=1, r^2=1, such that, the Illumina SNP rs6495306, included in our study is a perfect surrogate of rs588765. The SNPs are also equivalent in European Americans according to Hapmap project data (CEU: D'=1,

r^2=1; www.hapmap.org). We see no association of the SNP with alcohol dependence (OR=1.01, CI=0.95-1.08, p=0.7).

Association of rs1051730 with smoking quantity among subjects with alcohol dependence

A regression model was applied to examine the effect of the variant rs1051730 on smoking quantity within the alcohol dependence cohort. Significant association is seen between the variant and smoking quantity among subjects with an alcohol dependence diagnosis (p=0.002), with an estimated increase of 0.1 SQ levels per copy of the T allele of rs1051730, translating to approximately 1 cigarette per day (Table 2.4). This effect is the same in controls.

E. Discussion

A significant role for a common variant within the *CHRNA5/CHRNA3/CHRNB4* gene cluster has been firmly established by several studies linking the region to smoking-related phenotypes. These studies have examined smoking within the general population. It has been proposed that there are both common and specific genetic components to risk for different substance dependence disorders(54). Given that rs1051730 demonstrates a confirmed risk factor for one substance dependence, it is important to determine its role, if any, in dependence on other substances. This work has indeed already begun (77, 78). Our sample provides several strengths making it well-suited to this study. Our control group is very large, with over 30,000 subjects, and we have previously demonstrated significant association to nicotine dependence and smoking quantity in this group(67). Also, we have a single treatment center cohort which is the source for most of our cases, including all

illicit substance dependence cases. These cases have been phenotyped by treatment center physicians at the time of admission according to DSM-IV criteria.

We see no evidence of association of rs1051730, or any other SNP tested within the same LD block, with alcohol dependence, or dependence on opioids, amphetamines, cannabis, or SAH. The sample sizes available for study in substance dependence diagnoses other than alcohol are similar to those from other studies, but still too small to detect modest risks. We cannot therefore rule out a small effect of SNPs within this block in illicit substance dependence. However, providing all data here will allow for future meta-analyses to address this question with sufficient power. We could not replicate the reported association to alcohol dependence of another SNP within the LD block(77).

We do observe significant association of rs1051730 with smoking quantity within alcohol dependent subjects, with an effect size similar to that seen in the population controls. Our results suggest that the risk variant within the *CHRNA5*/*CHRNA3*/*CHRNB4* gene cluster which associates strongly with smoking and smoking-related diseases is a substance specific genetic risk factor for addiction.

Acknowledgements

We would like to thank all participants in the genetic studies whose contributions made this work possible, the staff at SAA Vogur Hospital, staff at Noatun (deCODE's recruitment center), and personnel at deCODE's core facilities. This work has been supported in part by funds from the European Commission (GENADDICT: LSHM-CT-2004-005166) and the National Institute of Drug Abuse (R01-DA017932)

Table 1. Demographic Description

Diagnosis	n	Males/Females	Age (years)	Comorbid Alcohol Dependence
Alcohol Dependence	2215	1409/806	50.1 ± 14.2	
Amphetamine Dependence	282	196/86	39.1 ± 9.4	271
Cannabis Dependence	374	279/95	37.5 ± 9.8	353
Opioid Dependence	196	109/87	45.9 ± 11.6	177
Sedative/Anxiolytic Dependence	357	184/173	48.9 ± 12.4	334
Population Controls	31 576	12 910/18 666	61.6 ± 21.6	

Table 2. Association of Variants in the LD Block with Alcohol Dependence

			Controls (n=31 576)	Alcohol Dependence (n=2215)		
SNP	Position (build 36)	Allele	Frequency	Frequency	OR (95% CI)	p-value
rs1394371	76511523	T	0.295	0.292	0.98 (0.91-1.06)	0.67
rs12903150	76511699	G	0.275	0.274	0.99 (0.92-1.07)	0.85
rs10519198	76529808	C	0.555	0.553	0.99 (0.93-1.06)	0.80
rs13180	76576542	T	0.615	0.604	0.95 (0.89-1.02)	0.17
rs3743079	76578115	T	0.168	0.173	1.04 (0.95-1.13)	0.44
rs8034191	76593077	T	0.658	0.664	1.03 (0.96-1.11)	0.43
rs3885951	76612971	T	0.916	0.916	1.00 (0.88-1.13)	0.99
rs2036534	76614002	T	0.772	0.769	0.98 (0.91-1.06)	0.66
rs6495306	76652947	G	0.433	0.436	1.01 (0.95-1.08)	0.73
rs680244	76658342	G	0.565	0.563	0.99 (0.93-1.06)	0.79
rs621849	76659915	G	0.435	0.437	1.01 (0.94-1.08)	0.82
rs1051730	76681393	T	0.340	0.335	0.98 (0.91-1.05)	0.55
rs8192475	76698284	G	0.949	0.951	1.02 (0.88-1.20)	0.76

SNPs are listed according to position in NCBI build 36. The reference allele, control frequency, affected frequency, odds ratio (OR) with 95% Confidence Interval (CI) and p-value are given for all SNPs tested within the LD block

Table 3. Association of Variants Tested to Other Substance Dependence Diagnoses

	Amphetamine Dependence (n=282)			Cannabis Dependence (n=374)			Opioid Dependence (n=196)			SAH Dependence (n=357)		
SNP	Frequency	OR (95% CI)	p-value	Frequency	OR (95% CI)	p-value	Frequency	OR (95% CI)	p-value	Frequency	OR (95% CI)	p-value
rs1394371	0.268	0.87 (0.72-1.06)	0.17	0.273	0.89 (0.76-1.06)	0.19	0.274	0.90 (0.72-1.13)	0.37	0.300	1.02 (0.86-1.21)	0.81
rs12903150	0.293	1.09 (0.90-1.32)	0.39	0.302	1.14 (0.97-1.34)	0.12	0.281	1.03 (0.82-1.28)	0.82	0.258	0.91 (0.77-1.09)	0.30
rs10519198	0.550	0.98 (0.82-1.16)	0.79	0.539	0.94 (0.80-1.09)	0.38	0.538	0.93 (0.76-1.14)	0.50	0.569	1.06 (0.91-1.23)	0.49
rs13180	0.594	0.91 (0.77-1.09)	0.32	0.608	0.97 (0.83-1.13)	0.70	0.599	0.94 (0.76-1.15)	0.53	0.591	0.90 (0.77-1.06)	0.20
rs3743079	0.156	0.92 (0.72-1.16)	0.47	0.159	0.94 (0.77-1.15)	0.54	0.176	1.06 (0.81-1.38)	0.67	0.169	1.01 (0.83-1.24)	0.91
rs8034191	0.688	1.15 (0.95-1.38)	0.15	0.671	1.06 (0.91-1.24)	0.46	0.676	1.09 (0.88-1.34)	0.45	0.651	0.97 (0.83-1.14)	0.72
rs3885951	0.908	0.90 (0.66-1.22)	0.49	0.913	0.95 (0.73-1.25)	0.73	0.911	0.93 (0.65-1.33)	0.70	0.920	1.05 (0.79-1.39)	0.73
rs2036534	0.748	0.88 (0.72-1.07)	0.20	0.759	0.93 (0.78-1.11)	0.42	0.770	0.99 (0.78-1.26)	0.93	0.756	0.92 (0.77-1.09)	0.33
rs6495306	0.433	1.00 (0.84-1.19)	0.99	0.425	0.97 (0.83-1.13)	0.68	0.441	1.04 (0.85-1.27)	0.74	0.408	0.90 (0.77-1.05)	0.19
rs680244	0.567	1.01 (0.85-1.20)	0.92	0.575	1.04 (0.89-1.21)	0.61	0.556	0.96 (0.79-1.18)	0.72	0.592	1.12 (0.96-1.30)	0.16
rs621849	0.433	0.99 (0.83-1.18)	0.91	0.425	0.96 (0.83-1.12)	0.60	0.444	1.04 (0.85-1.27)	0.73	0.408	0.89 (0.77-1.04)	0.15
rs1051730	0.316	0.89 (0.74-1.08)	0.23	0.332	0.96 (0.82-1.13)	0.63	0.327	0.94 (0.76-1.16)	0.57	0.346	1.03 (0.87-1.20)	0.76
rs8192475	0.952	1.06 (0.71-1.58)	0.77	0.951	1.02 (0.73-1.45)	0.89	0.934	0.75 (0.49-1.14)	0.18	0.951	1.03 (0.73-1.47)	0.85

All p values shown are two sided. For each substance dependence diagnosis the n, frequency, OR with 95% CI and p-value are shown for all variants tested within the LD block. The reference allele is that shown in Table 2, and all tests use the control values shown in Table 2.

Table 4. Regression Analysis of rs1051730 and Smoking Quantity

	Substance Dependence (n=1691)		Controls (n=11 241)	
	Estimate (95% CI)	P	Estimate (95% CI)	P
Copies of T allele	0.100 (0.068-0.132)	0.0019	0.101 (0.089-0.113)	$2x10^{-18}$
sex (male)	0.343 (0.339-0.387)	$1x10^{-14}$	0.395 (0.378-0.411)	$<10^{-20}$
year of birth (categorical)	-	$< 10^{-10}$	-	$<10^{-9}$
allele x sex	-	0.78	-	0.33
allele x age	-	0.63	-	0.80

Multiple regression of SQ level on allele T, sex and year of birth, giving adjusted values for each explanatory variable adjusting for the others. For the tests of interaction, the interaction terms involving the variant were individually added to the initial model.

Supplementary Table

A. Alcohol and Amphetamine Dependence

			Controls (n=31 576)	Alcohol Dependence (n=2215)			Amphetamine Dependence (n=282)		
SNP	Postion (build 36)	Allele	Frequency	Frequency	OR (95% CI)	p-value	Frequency	OR (95% CI)	p-value
rs7171749	76495769	T	0.635	0.650	1.07 (1.00-1.15)	0.06	0.651	1.07 (0.90-1.29)	0.44
rs4887053	76499753	C	0.795	0.787	0.96 (0.88-1.04)	0.28	0.766	0.85 (0.69-1.04)	0.11
rs6495309	76702299	T	0.214	0.216	1.02 (0.94-1.10)	0.68	0.238	1.15 (0.93-1.41)	0.19
rs1948	76704453	T	0.342	0.353	1.05 (0.98-1.13)	0.17	0.355	1.06 (0.88-1.27)	0.54
rs950776	76713072	T	0.653	0.650	0.99 (0.92-1.06)	0.71	0.626	0.89 (0.74-1.07)	0.21
rs12594247	76733687	T	0.152	0.156	1.04 (0.95-1.14)	0.43	0.165	1.11 (0.87-1.40)	0.40
rs12900519	76736181	T	0.932	0.931	0.98 (0.86-1.12)	0.78	0.936	1.07 (0.75-1.52)	0.71
rs1996371	76743860	G	0.391	0.386	0.98 (0.91-1.05)	0.53	0.351	0.84 (0.70-1.01)	0.06
rs6495314	76747583	C	0.390	0.383	0.97 (0.91-1.04)	0.39	0.349	0.84 (0.70-1.00)	0.06
rs8032156	76751552	G	0.838	0.829	0.94 (0.86-1.03)	0.17	0.830	0.95 (0.75-1.20)	0.65
rs8038920	76761599	G	0.623	0.624	1.00 (0.93-1.07)	0.95	0.599	0.90 (0.76-1.08)	0.27
rs4887077	76765418	T	0.385	0.380	0.98 (0.92-1.05)	0.61	0.340	0.83 (0.69-0.99)	0.04
rs11638372	76770613	T	0.388	0.383	0.98 (0.92-1.05)	0.59	0.348	0.84 (0.70-1.01)	0.06
rs11072793	76793496	G	0.233	0.237	1.02 (0.95-1.11)	0.55	0.253	1.11 (0.91-1.36)	0.30
rs2277547	76869485	T	0.749	0.744	0.97 (0.90-1.05)	0.50	0.738	0.94 (0.77-1.15)	0.57
rs3743057	76876061	G	0.722	0.724	1.01 (0.94-1.09)	0.80	0.700	0.90 (0.75-1.09)	0.29

The first 2 SNPs are located upstream of the 5' end of the LD block, while the remaining SNPs are located downstream of the 3' end.

B. Cannabis, Opioid and Sedative, Anxiolytic or hypnotic (SAH) Dependence

	Controls (n=31 576)	Cannabis Dependence (n=374)			Opioid Dependence (n=196)			SAH Dependence (n=357)		
SNP	Frequency	Frequency	OR (95% CI)	p-value	Frequency	OR (95% CI)	p-value	Frequency	OR (95% CI)	p-value
rs7171749	0.635	0.643	1.04 (0.89-1.21)	0.64	0.661	1.12 (0.91-1.39)	0.29	0.679	1.22 (1.04-1.43)	0.02
rs4887053	0.795	0.778	0.91 (0.76-1.09)	0.29	0.793	0.99 (0.77-1.27)	0.95	0.779	0.91 (0.76-1.09)	0.31
rs6495309	0.214	0.233	1.12 (0.93-1.34)	0.23	0.219	1.03 (0.81-1.32)	0.78	0.235	1.13 (0.95-1.36)	0.18
rs1948	0.342	0.349	1.03 (0.88-1.21)	0.69	0.344	1.01 (0.82-1.25)	0.91	0.319	0.90 (0.77-1.06)	0.22
rs950776	0.653	0.639	0.94 (0.81-1.1)	0.45	0.656	1.01 (0.82-1.25)	0.91	0.687	1.17 (0.99-1.37)	0.06
rs12594247	0.152	0.156	1.04 (0.84-1.28)	0.72	0.161	1.07 (0.81-1.41)	0.62	0.186	1.28 (1.05-1.57)	0.02
rs12900519	0.932	0.945	1.26 (0.91-1.73)	0.16	0.934	1.03 (0.69-1.53)	0.90	0.926	0.91 (0.68-1.22)	0.52
rs1996371	0.391	0.381	0.96 (0.82-1.12)	0.57	0.388	0.99 (0.80-1.21)	0.89	0.370	0.91 (0.78-1.07)	0.25
rs6495314	0.390	0.378	0.95 (0.82-1.11)	0.52	0.380	0.96 (0.78-1.18)	0.69	0.366	0.90 (0.77-1.05)	0.19
rs8032156	0.838	0.842	1.03 (0.84-1.27)	0.75	0.826	0.92 (0.70-1.20)	0.53	0.806	0.81 (0.66-0.98)	0.03
rs8038920	0.623	0.607	0.93 (0.80-1.09)	0.38	0.633	1.04 (0.85-1.28)	0.70	0.644	1.09 (0.93-1.28)	0.26
rs4887077	0.385	0.370	0.94 (0.81-1.10)	0.44	0.375	0.96 (0.78-1.18)	0.70	0.359	0.89 (0.76-1.05)	0.16
rs11638372	0.388	0.376	0.95 (0.81-1.11)	0.52	0.375	0.95 (0.77-1.17)	0.61	0.364	0.90 (0.77-1.06)	0.21
rs11072793	0.233	0.234	1.00 (0.84-1.20)	0.96	0.250	1.10 (0.87-1.39)	0.44	0.279	1.27 (1.07-1.51)	0.007
rs2277547	0.749	0.754	1.03 (0.87-1.22)	0.74	0.740	0.95 (0.76-1.20)	0.69	0.718	0.86 (0.72-1.02)	0.08
rs3743057	0.722	0.721	0.99 (0.84-1.18)	0.95	0.719	0.99 (0.79-1.24)	0.92	0.725	1.02 (0.86-1.21)	0.82

Chapter 3

Characterization of the *CHRNA5/CHRNA3/CHRNB4* nicotinic acetylcholine receptor gene cluster in nicotine dependence and lung cancer

Material from this chapter also appears in an article which will be submitted to Human Molecular Genetics. Contributing authors of this article will include: Anna Wiste, Frank Geller, Gudrun A. Jonsdottir, Aslaug Jonasdottir, Adalbjorg Jonasdottir, Jelena Kostic, Thorunn Rafnar, Kari Stefansson, Thorgeir E. Thorgeirsson

A. Abstract

Recent genome-wide association scans of smoking quantity, nicotine dependence and lung cancer have all highlighted an important role of sequence variants within the *CHRNA5*/*CHRNA3*/*CHRNB4* gene cluster on chromosome 15. In order to search for variants responsible for these association signals, we have sequenced exons and flanking sequences of all 3 genes in 184 lung cancer cases, 176 nicotine dependent cases, and 175 low-quantity smokers. A total of 111 variants were identified of which 47 are novel. Variants of particular interest because of functional importance or observed association were carried into additional genotyping. Included among these is rs3841324, a common insertion/deletion in the promoter of *CHRNA5*, which is proposed to influence expression of the gene. We find evidence of strong association of this variant to expression of the gene in blood ($r=0.72$, $p=4x10^{-71}$) and adipose tissue($r=0.73$, $p=2x10^{-63}$). No variants demonstrated stronger association with lung cancer or nicotine dependence than rs1051730, previously reported in our genome-wide scan of smoking quantity, and also associated with lung cancer and nicotine dependence. This SNP is equivalent in Iceland to the non-synonymous coding SNP rs16969968 located in *CHRNA5*. Our data suggest that this functional polymorphism remains the most likely cause of the association with smoking and smoking-related diseases.

B. Introduction

The *CHRNA5*/*CHRNA3*/*CHRNB4* gene cluster on chromosome 15 has recently received much attention due to associations with smoking behavior and smoking-related diseases. We identifed the T allele of the synonymous SNP rs1051730 as a risk factor for increased daily cigarette consumption in a large genome-wide association study(67), and showed that the allele is also significantly associated with nicotine dependence, lung cancer and peripheral arterial disease. Other studies have also provided evidence of association with smoking quantity(64) and nicotine dependence(38), and two genome-wide association studies also identified the same variant as a risk factor for lung cancer(65, 66).

The presence of association signals in genome-wide scans and tagging SNP approaches does not necessarily identify the actual variants responsible for association, because the observed associations can be indirect through linkage disequilibrium with the relevant functional variants. The SNP highlighted in most studies, rs1051730, is a synonymous coding SNP in *CHRNA3*. Another sequence variant, the non-synonymous coding SNP rs16969968 in *CHRNA5*, has also been reported(38), and the two variants are in strong linkage disequilibrium in populations of European ancestry (D'=1, r^2=0.9; www.hapmap.org).

Initial characterizations of the cluster focused on the genes as candidates for rare mendelian disorders, including megacystitis-microcolon-hypoperistalsis syndrome (*CHRNA3*/*CHRNB4*)(79) and autosomal dominant nocturnal frontal lobe epilepsy (*CHRNA5*/*CHRNA3*/*CHRNB4*)(80, 81). A recent candidate gene study of the role of nicotinic acetylcholine receptor genes in smoking behavior used resequencing as an

approach to identify tagging SNPs for genotyping in a larger cohort(42). In this study groups of 72 low and high quantity smokers were used in addition to 40 non-smoking controls. Their approach provided further confirmation of rs1051730 as a risk factor for increased smoking quantity. It is not yet clear which polymorphism is responsible for the associations seen with smoking behavior and smoking-related diseases. We have therefore resequenced all three genes using three independent groups: lung cancer cases, nicotine dependent individuals, and low quantity smokers. We have examined the potential role of both common and rare polymorphisms in these groups, taking into account the known association with the SNP rs1051730.

C. Results

All exons, promoters, and 5' and 3'UTRs were sequenced from each of the three genes in this nicotinic acetylcholine receptor subunit cluster in a sample of lung cancer patients (n=184), nicotine dependent smokers (n=176) and low quantity smokers (n=175). In total, 111 variants were found, 47 of which were not present in dbSNP129. A full description of all variants is found in Table 3.7, including position, alleles, frequency and possible functional significance. Statistical analysis focused on 50 variants with minor allele frequencies greater than 1%. Results of this analysis are found in Table 3.1. Given the strong established effect seen with rs1051730, we expect to find significant results for this SNP and correlated SNPs. P-values which include an adjustment for the effect of rs1051730 are thus also included in the table.

We examined linkage disequilibrium (LD) among these polymorphisms in order to define equivalence groups in which all polymorphisms have $r^2>0.8$ to one SNP identified as head of the group in Table 3.2. Six equivalence groups are formed accounting for all but three of the polymorphisms with frequency greater than 5% (See Table 3,2). These three polymorphisms had strongest LD to the head of class A (rs1051730; r^2 between 0.64 and 0.79) and are thus reported together with that group in Table 3.2.
Genotypes from Illumina Human Hap300 chips are available for all subjects sequenced, as well as for additional subjects in each group. Information on linkage disequilibrium within the sequencing sample was used to identify appropriate tagging variants from the Illumina chip to effectively increase sample size for variants of interest.

Variants of Interest

rs16969968

The non-synonymous *CHRNA5* variant rs16969968 has been highlighted in the literature previously (38). LD in European Americans is strong between this variant and rs1051730 according to the Hapmap project data (D'=1, r^2=0.9; www.hapmap.org). We found the 2 variants to be equivalent in our sequencing sample.

rs1051730 equivalence group

In addition to rs16969968, several other SNPs were found to be in very strong LD with rs1051730 in Iceland. These include rs55853698, rs55781567 and rs8192482, all with r^2>0.93 to rs1051730/rs16969968. Because LD is so strong in Iceland, we cannot differentiate between these 5 SNPs. Another SNP, ss107794645, exhibited weaker LD with rs1051730/rs16969968 (D'=0.91, r^2=0.69). Within the sequencing sample this SNP gave a stronger risk than rs1051730 for nicotine dependence (OR=1.65 vs. 1.49) but not lung cancer (OR=1.53 vs 1.58). A single SNP assay (Centaurus; Nanogen) was designed to further test this variant in Iceland. After additional subjects were genotyped, the OR of this variant is 1.26 (p=0.006, p=0.8 after adjustment for rs1051730) for lung cancer (n=645), and is 1.18 (p=0.02, p=0.8 after adjustment for rs1051730) for nicotine dependence (n=2068), both tested against low quantity smokers (n=535),. These results indicate that risk associated with ss107794645 is due to LD with rs1051730.

rs12907519/rs8192475

The results from our sequencing analysis alone indicate a significant protective effect of the C allele of rs12907519, a SNP located in intron 1 of *CHRNA3*. With low quantity smokers as controls, the variant has an OR of 0.34 for nicotine dependence

(p=0.007 after adjustment) and 0.21 for lung cancer (p=0.0003 after adjustment). This SNP is within equivalence group D, in strong LD with rs8192475 (r^2=0.93) which is included on the Illumina chips. With all genotypes available for rs8192475, association of this variant is not significant for lung cancer (OR=0.78, p=0.5 after adjustment for rs1051730) or nicotine dependence (OR=0.87, p=0.9 after adjustment) when compared to low quantity smokers (see Table 3.3). Given the strong LD between these variants, we can rule out association of rs12907519 with either lung cancer or nicotine dependence.

Equivalence Classes in Illumina samples

Four equivalence classes are headed by a SNP on the Illumina chip. A fifth can be tagged with r^2=0.98 by a haplotype of two SNPs from the chip. Results within the larger chip sample are displayed for all tagged classes in Table 3.3, with and without adjustment for the effect of rs1051730. One class (A) is headed by rs1051730. Within the chip genotyped sample analyzed here, the T allele is strongly associated with both nicotine dependence (OR=1.4, p=$7.4\text{x}10^{-15}$) and lung cancer (OR=1.52, p=$1.5\text{x}10^{-11}$). Of the SNPs which head the remaining 4 classes tagged by Illumina chips, with and without correction for rs1051730, only rs8192475 displayed significant association in any of the three tests within the sequencing sample. In the larger chip-genotyped sample, several SNPs have significant p-values due to correlation with rs1051730. After adjustment for the effect of rs1051730 the SNP rs1948 has a p-value of 0.006. This presents the possibility that a protective effect for lung cancer might exist for a variant in this equivalence class which occurs primarily on the same background as the risk effect of rs1051730. Any such effect would be small, and is masked by the comparably strong risk associated with rs1051730.

rs578776

The SNP rs578776 has recently been reported to be an independent, second risk variant for nicotine dependence within this LD block(75). We genotyped additional nicotine dependent cases and low quantity smokers so that our data set would be large enough to address the relationship of rs578776 to rs1051730/rs16969968. According to Hapmap project data, in European Americans LD between the variants is D'=1, r^2=0.2 (www.hapmap.org). In Iceland we see similar results (D'= 0.99, r^2=0.19, n=3026). The risk allele of rs1051730/rs16969968 is fixated on the background of the major allele of rs578776. Therefore there are only 3 haplotypes possible. We find that all the risk associated with rs578776 is confined to the haplotype which includes the risk variant of rs1051730/rs16969968 (OR=1.34, p=$1.56x10^{-4}$; See Figure 3.1). The frequency of the haplotype containing the protective allele of rs1051730 and the risk allele of rs578776 occurs at a lower frequency in nicotine dependence (37.6%) compared to low quantity smokers (39.7%). While this study provides no evidence to support an independent risk for nicotine dependence associated with this variant, a larger sample size is needed to exclude this possibility.

Rare Variants

Of the variants identified with sequencing, 59 occur at frequencies of less than 1%. Table 3.8 includes the number of carriers in each phenotype group for each of these variants. Among them are 7 missense mutations and one 20bp exonic deletion. The exonic deletion occurs in *CHRNA3* in one subject from the nicotine dependence group. This individual received a score of 4 on the FTND scale and did not meet DSM criteria for nicotine dependence. None of the rare variants alone can fully account for the signal

observed. However, we cannot rule out the possibility that among these variants are rare high penetrance variants which might influence risk of one or both conditions. Testing for association with a larger sample size is needed to evaluate these SNPs.

Three length polymorphisms: rs3841324, rs55787222, and rs60706203

Three length polymorphisms were genotyped directly in additional subjects. These include a 22bp insertion/deletion, rs3841324, in the promoter of *CHRNA5*, identified in a scan for promoter polymorphisms affecting gene expression(99), rs60706203, a 3bp insertion/deletion in the leader sequence of *CHRNA3*, and rs55787222, a 4bp microsatellite in the promoter region of *CHRNA3*. Additional rare alleles of each of the last 2 variants were identified with further genotyping (see Table 3.7).

Results for association analysis of these markers are presented in Table 3.4. P-values were adjusted to take into account the effect of rs1051730. There is no significant association with lung cancer or nicotine dependence for either rs3841324 or rs60706203. In the case of rs55787222, a 4 bp microsatellite in the promoter region of *CHRNA3*, the allele containing 2 copies of the 4bp sequence is not associated with either condition before correction for rs1051730. After correction, however, the p- value is 0.004 for association with nicotine dependence. Within this sample, the p-value for rs1051730 is 0.001 for nicotine dependence. It appears that the allele of rs55787222 which contains 2 copies may be protective against nicotine dependence. The risk allele of rs1051730 is fixated on this background, and the risk contributed by this variant is stronger than the protective effect which may be supplied by this allele of rs55787222. However, the risk for rs1051730 is observed for the comparison of both nicotine dependence and lung

cancer against low quantity smokers. The possible protective effect of rs55787222 is only observed for nicotine dependence.

Expression

We measured expression of *CHRNA5* in two tissues to address whether genetic variants in the cluster are associated with expression regulation. In particular, rs3841324 has been reported as a promoter regulatory element in cell culture(99). We sought to test the effect of this variant on expression *in vivo*. Expression of *CHRNA5* was strongly associated with rs3841324 genotype, with relative expression levels higher for the short allele in blood ($r=0.72$, $p=4x10^{-71}$) and subcutaneous adipose tissue ($r=0.73$, $p=2x10^{-63}$; See Figure 3.2). Association with expression of *CHRNA5* was also examined for the other SNPs within the LD block, with one marker from each equivalence class tested (see Table 3.5). All markers were significantly associated with expression. Adjusting for rs3841324 reduces the significance of the association for the other SNPs drastically, in subcutaneous adipose tissue only rs1051730 remains nominally significant ($p=0.018$) and three SNPs show nominally significant association in blood (minimum $p=0.006$ for rs1051730). Overall, expression in blood and subcutaneous adipose tissue is strongly associated with rs3841324. However, we cannot rule out an additional comparably weak effect of another SNP, which was best captured by rs1051730. Expression of *CHRNA5* was not associated with lifetime regular smoking, or with smoking within the past 24 hours (data not shown).

We have established that there is no risk for nicotine dependence or lung cancer associated with this variant independent of the risk associated with rs1051730 (See Table 3.4). However, there is strong LD between the two variants. The T allele of rs1051730

only appears on the haplotype background including the long, or low expression, allele of rs3841324.

A careful characterization of the CHRNA5/CHRNA3/CHRNB4 cluster does not identify any variants with stronger association to nicotine dependence or lung cancer than rs1051730/rs16969968. Therefore the SNP non-synonymous SNP rs16969968 remains the variant most likely to have functional effects leading to the observed association signals within this region.

D. Discussion

A number of recent reports implicating the nicotinic acetylcholine receptor gene cluster on chromosome 15q25 in both smoking and lung cancer have focused attention of researchers on the region. These studies have employed genome-wide association analysis with chip genotyping or tagging SNPs in candidate gene studies(38, 42, 64-67). A more thorough investigation of genetic variation within this region is an important step in helping to determine the functional variant responsible for the observed association signals. We set out to sequence the region using samples large enough to also identify additional variants with significant effects in one or both conditions. The choice of low quantity smokers as a control group is guided by our previous finding that the variant is not associated with smoking initiation but rather with smoking quantity. The frequency of the variant is significantly lower in low quantity smokers compared with population controls. Affected groups include both nicotine dependent smokers and lung cancer patients, allowing for the possibility of differences between these groups as well. The effect of rs1051730 on smoking behavior as it has been measured in studies so far is not

strong enough to fully account for the effect of the variant on lung cancer, as has been stated by us(67), and others(76). Presence of additional correlated variants in the region with risk specific for lung cancer but no effect on smoking behavior could account for this observed difference.

The primary result of our study is that no variant within the regions sequenced displays a stronger association with either nicotine dependence or lung cancer than that seen for the T allele of rs1051730. We have determined that in Iceland, this variant is equivalent to rs16969968, a non-synonymous coding SNP in *CHRNA5*, previously highlighted in a candidate gene study(38). The risk allele of this variant, A, codes for an asparagine at amino acid position 398 while the protective allele, G, codes for arginine. The lack of any other functional variants within this region showing strong assoication with lung cancer or nicotine dependence suggests that rs16969968 is the variant most likely explaining the results.

We do not see any variants with significant differences in frequencies between lung cancer cases and nicotine dependent smokers. However, there are many rare variants within the region and some have higher frequencies in the lung cancer cases. We have provided all of the data obtained on these rare variants, so that they may be utilized in future studies. It remains possible that there are rare variants within this region which contribute to lung cancer risk or to nicotine dependence risk. None of the rare variants identified in our study could, alone, account for the association seen with this variant.

A large number of variants are tested between three groups making interpretation of the significance of the findings difficult. Testing results of interest in additional groups can aid correct interpretation of findings. We have available a large number of individuals genotyped on Illumina chips and the possibility of also genotyping additional

markers on these individuals. The observed association of rs12907519 with nicotine dependence and lung cancer after correction for rs1051730 in the sequencing group is a good example of a relatively strong effect identified in a small group which cannot be confirmed in the larger study group available.

Our characterization of the region also included examination of expression regulation, targeting in particular the insertion/deletion in the promoter of *CHRNA5*, rs3841324. This variant was found to affect transcription in a large scan of promoter elements(99). The ability of each promoter sequence to promote luciferase transcription was tested in human cell lines. The insertion/deletion rs3841324 was found to have a significant effect on luciferase transcription, and analysis of the sequence indicated that transcription factor binding sites for SP1 and AP2 are altered by this variant. Genetic regulation of expression of *CHRNA5* was also addressed in a recent association analysis of this region in alcohol dependence(77). This report focused on two variants, rs3841324 and a SNP rs588765, with expression of *CHRNA5* measured in brain tissue from a small sample of subjects (n=43). Significant association was found with both variants, and the authors concluded that it could not be determined which variant was responsible for the association due to linkage disequilibrium. The SNP rs588765 is an intronic SNP and thus could not account for the expression regulation seen in the Buckland et al. study, which specifically tested promoter sequences. Our study has used a large cohort of individuals with samples from two tissues, blood and subcutaneous adipose tissue, collected specifically for studies of genetic regulation of gene expression(100). The effect size of rs3841324 is large and shows remarkable concordance between the two different tissues tested, indicating that this variant provides strong, consistent regulation of transcription of *CHRNA5 in vivo*. Genotyping rs3841324 in the larger sample of cases indicated that with

correction for rs1051730 there is no association of this variant with either nicotine dependence or lung cancer. However, this does not mean that it does not have an important role to play in the study of this cluster in smoking behavior and lung cancer. In Iceland the risk allele of rs1051730/rs16969968 only appears on a background containing the low expression allele of *CHRNA5*. It will be relevant for future *in vivo* and *in vitro* studies to recognize that the resulting amino acid substitution caused by rs16969968 is only present in *CHRNA5* receptors subunits transcripts which are expressed at low levels.

A key point here is that the characterization of this risk variant has thus far only been performed in individuals of European ancestry. Data from the Hapmap project indicate a significant ethnic difference in the frequency of the variant rs16969968, with the risk allele present at much higher frequency in individuals of European ancestry than in African and Asian populations. Phase 3 Hapmap project data for this SNP indicates a frequency of approximately 40% in Europeans, 10% in Maasai, and less than 5% in East Asian populations (www.hapmap.org). The effect of the variant may be significantly different in these populations. Furthermore, possible interactions with other functional variants in the region may be different in populations where the linkage disequilibrium structure differs considerably.

E. Materials and Methods

Subjects for Sequencing

Three groups of subjects were selected for sequencing analysis: (1) lung cancer patients (n=184), (2) nicotine dependent smokers without other addictions (n=176) and (3) low-quantity smokers (n=175) (See Table 3.6 for demographic information). Low-

quantity smokers reported regular smoking for at least one year and reported only social smoking or less than 5 cigarettes per day. Subjects with lung cancer show the highest frequency of the identified risk variant, and generally constitute a population with high lifetime smoking exposure. Our sample of nicotine dependent individuals received the diagnosis based on questionnaire data addressing two systems of classification of nicotine dependence, the Fagerstrom Test for Nicotine Dependence (FTND)(35) and the criteria of the Diagnositic and Statistical Manual, Version IV (DSM). Subjects met criteria under either or both systems (FTND 4+ or DSM 3+). Individuals with other substance dependence or abuse diagnoses were excluded. Our previous analysis indicated that the effect of the risk variant was to increase smoking quantity among smokers, rather than affecting initiation. Therefore, we used smokers with low consumption as a control group for study.

Subjects for additional genotyping

Certain variants of interest were specifically genotyped in additional individuals. For the length variants rs55787222, rs3841324 and rs60706203, the subjects included 567 lung cancer patients, 1623 nicotine dependent smokers and 608 low quantity smokers (See Table 3.9).

All subjects sequenced have also been genotyped with Human Hap300 or Human Hap300-duo1 Bead Arrays (Illumina; San Diego, CA, USA). Additional subjects from each group have also been genotyped using these chips. LD information obtained in the sequencing cohort was used to identify tagging SNPs for testing in the larger sample, which included 669 lung cancer cases, 1950 nicotine dependent smokers and 4680 low quantity smokers (See Table 3.9)

The study protocols were approved by the National Bioethics Committee (NBC) and the Data Protection Authority (DPA) of Iceland. The DPA has encrypted all personal identifiers linked to phenotype information or blood samples using a third-party encryption system(95). All subjects are of Icelandic ancestry.

Sequencing

The exons, 5' and 3' UTRs, and flanking sequences 1kb upstream of *CHRNA5*, *CHRNA3*, and *CHRNB4* were sequenced. Sequence for the region was obtained from NCBI build 36. A total of 57 primer pairs were designed (see Table 3.10). The position of regions sequenced (build 36) can be found in Table 3.11. PCR amplification and sequencing reactions were set up on Zymark ALH300 workstations, with amplification performed on MJR Tetrads. PCR products were purified using AMPure (Agencourt Bioscience). Dye terminator removal was performed using CleanSEQ (Agencourt) to repurify. Electrophoresis was performed on Applied Biosystems 3730 DNA Analyzers. Sequence editing and analysis were performed using deCODE Genetics Sequence Miner software. SNP calling was done by both manual inspection and automated calling. All SNPs identified through automated calling were then confirmed by manual inspection of the sequence traces. Insertion/deletions and microsatellites were identified by manual inspection. Simple, rare insertion/deletions were called manually.

Genotyping

Additional genotyping of SNPs was done using the Centaurus platform (Nanogen). Three variants, rs55787222, rs3841324, and rs60706203, observed in the

sequencing, were genotyped in a larger population. For these markers primers were designed using Primer3. PCR reactions were set up on Zymark ALH300 workstations and amplification performed on MJR Tetrads. PCR products were pooled, an internal size standard added, and then resolved on Applied Biosystems 3730 DNA Analyzers. Primers and PCR conditions are available on request. Genotypes were called and edited using deCODE Allele Caller and deCODE-GT.

Expression Analysis

The variant rs3841324 was identifed as a promoter element with significant effect on transcription of *CHRNA5* in a genome scan for regulatory elements(99). We therefore examined its role in regulating expression of the gene in blood and subcutaneous adipose tissue using an expression cohort previously described(100). From this cohort, genotype and expression data were used from 446 individuals with blood samples and 376 individuals with subcutaneous adipose tissue samples.

RNA samples were purified using RNeasy Mini Kit (Quiagen), and integrity analyzed using Agilent 2100 Bioanalyzer. Total RNA was converted to cDNA using the High Capacity cDNA Archive Kit (Applied Biosystems). Two Taqman assays were designed for *CHRNA5*, so that positive results cannot be attributable to the specific assay used. The probes are located at different exon boundaries, one crossing exon 2 and 3, and the other crossing exons 3 and 4. Real-time PCR was carried out according to manufacturer's recommendations on an ABI Prism 7900HT Sequence Detection System. Quantification was performed using the ΔΔCt method (User Bulletin no. 2, Applied Biosystems 2001). A housekeeping gene, in this case *GUSB*, was run in parallel for normalization.

Statistical Analysis

A likelihood ratio test was used for analysis using χ^2 statistics. In all cases p-values are reported both with and without correction for the effect of rs1051730. P-values are reported without correction for multiple testing. In the analysis of the larger samples generated from Illumina genotypes and individual genotyping of length polymorphisms, p-values are corrected for relatedness among affecteds as described previously using a simulation procedure with the known genealogy(25).

The expression data were log-transformed, adjusted for sex and age with a linear regression model, and the standardized residuals were used as the variable. There were 307 individuals present in both data sets and their residuals for the two tissues tested were highly correlated ($r=0.65$, $p=7x10^{-39}$).

In analysis of equivalence classes in larger cohorts, genotypes for rs569207 are inferred. Allele T is tagged by a haplotype of allele C at rs1051730 and allele G at rs680244 ($r^2=0.98$ in the sequencing data) in the analysis of Illumina data. In the expression analysis genotypes were inferred using a two SNP haplotype based on allele G at rs680244 and allele T at rs578776 ($r^2=0.99$ in the sequencing data).

Figure 3.1: Illustration of the risk for nicotine dependence observed for rs1051730 (1) and rs578776 (2) based on the comparison of 2161 nicotine dependent individuals and 865 low quantity smokers. The frequencies for cases and controls are given in parentheses below the alleles (haplotypes), the arrows point towards the allele (haplotype), for which the risk is observed. (a) displays the odds ratios observed for the two SNPs and the linkage disequilibrium between them, (b) shows the odds ratios between the three observed haplotypes. There is no significant odds ratio for the haplotype with the protective allele at rs1051730 and the risk allele at rs578776 compared with the haplotype with both protective alleles. The comparison of the haplotype with the protective allele at rs1051730 and the risk allele at rs578776 against the haplotype with both high risk alleles shows a significant odds ratio due to rs1051730 allele T.

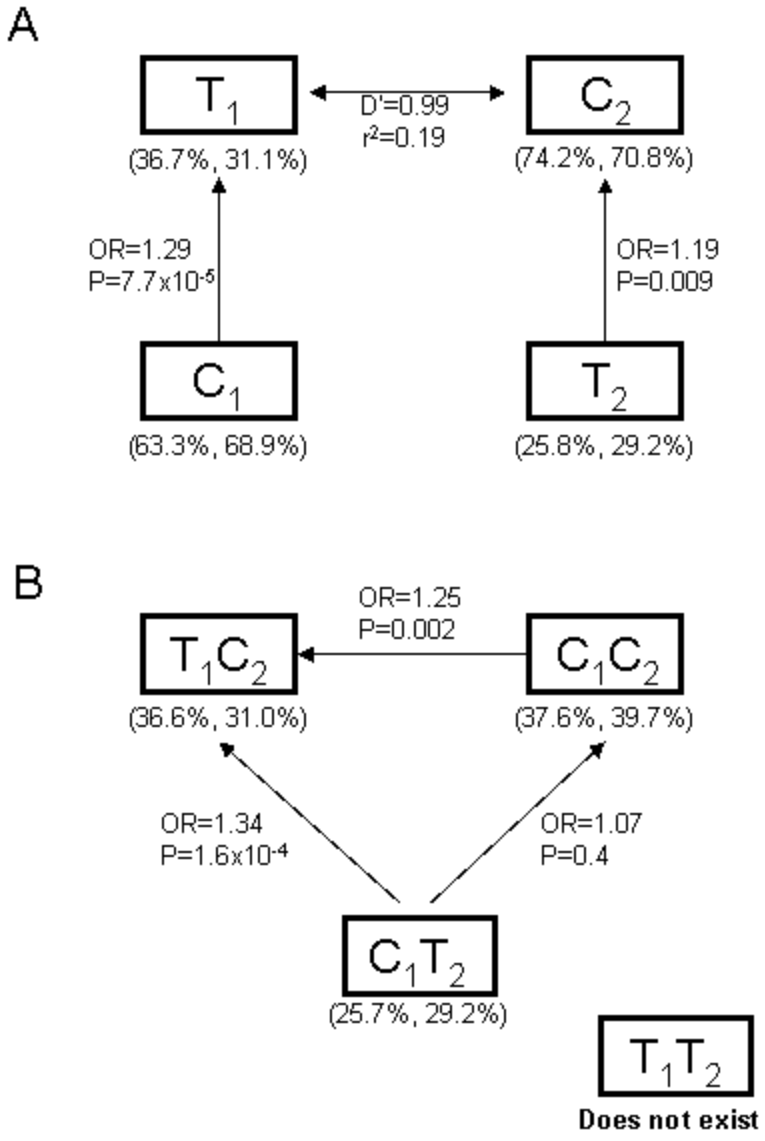

Figure 3.2 – The short allele of rs3841324 is associated with increased expression of CHRNA5 RNA in blood and subcutaneous adipose tissue. Expression of *CHRNA5* for the three different genotypes of rs3841324 was measured with real-time PCR using cDNA derived from RNA for blood (n=446) and subcutaneous adipose tissue (n=376). The expression of *CHRNA5* is measured relative to a housekeeping gene *GUSB*. The error bars indicate the standard error of the mean. Log-transformed relative expression values were adjusted for age and sex using a linear regression model. Correlation of the standardized residuals of the model yields a highly significant association with rs3841324 genotype in both blood ($r=0.72$, $p=4x10^{-71}$) and subcutaneous adipose tissue ($r=0.73$, $p=2x10^{-63}$).

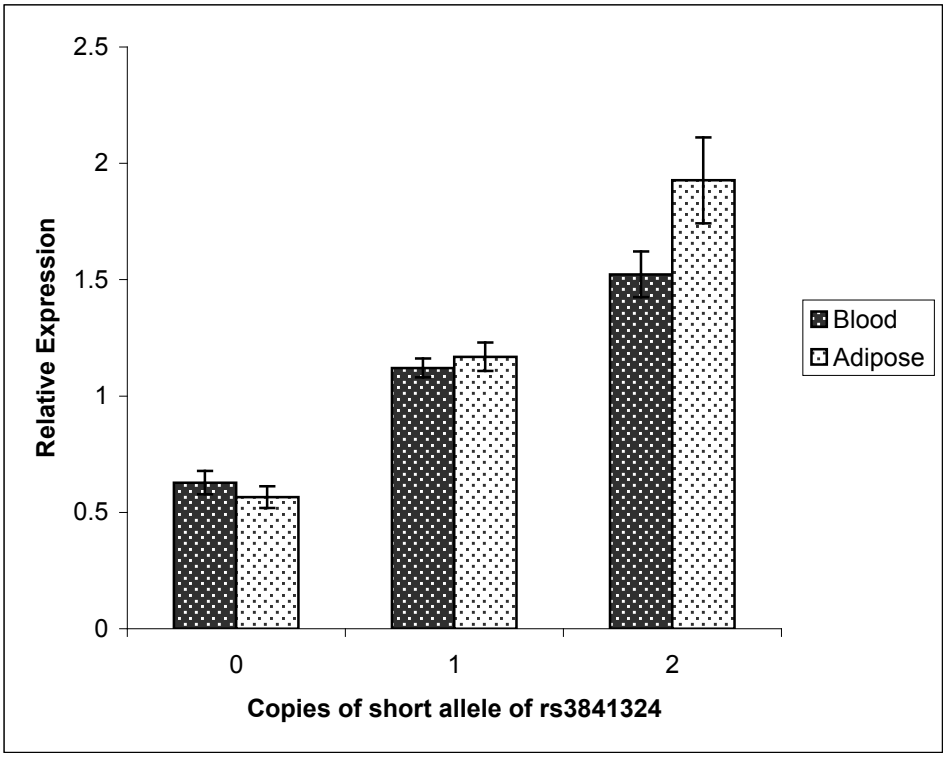

Table 3.1 - Comparison of frequency of markers with minor allele frequency >1% in Lung Cancer (LC), Nicotine Dependence (ND) and Low Quantity Smokers (LQS).

		ND		LQS		LC		ND against LQS			LC against LQS			LC against ND		
Marker	**allele**	**N**	**Freq**	**N**	**Freq**	**N**	**Freq**	**OR**	**P**	**Padj**	**OR**	**P**	**Padj**	**OR**	**P**	**Padj**
rs1051730	**T**	**176**	**0.384**	**175**	**0.294**	**184**	**0.397**	**1.49**	**0.01**	**-**	**1.58**	**0.004**	**-**	**1.06**	**0.7**	**-**
rs12898919	C	175	0.040	172	0.076	178	0.028	0.51	0.04	0.1	0.35	0.004	0.01	0.69	0.4	0.4
rs12899226	G	174	0.040	173	0.081	182	0.027	0.48	0.02	0.06	0.32	0.001	0.005	0.67	0.3	0.4
rs12907519	C	160	0.034	158	0.095	162	0.022	0.34	0.002	0.007	0.21	0.00004	0.0003	0.62	0.3	0.3
rs12911814	C	174	0.040	172	0.084	172	0.029	0.46	0.02	0.04	0.33	0.001	0.006	0.71	0.4	0.4
rs12914008	T	175	0.031	171	0.056	182	0.019	0.55	0.1	0.2	0.33	0.009	0.02	0.60	0.3	0.3
rs16969968	A	171	0.389	174	0.293	184	0.397	1.53	0.008	1.0	1.59	0.004	1.0	1.03	0.8	1.0
rs1948	T	176	0.335	174	0.351	184	0.351	0.93	0.7	0.3	1.00	1.0	0.09	1.07	0.7	0.4
rs2229961	A	176	0.017	174	0.003	183	0.014	6.02	0.05	0.1	4.81	0.1	0.2	0.80	0.7	0.7
rs28534575	C	172	0.183	170	0.244	182	0.206	0.69	0.05	0.3	0.80	0.2	0.9	1.16	0.4	0.3
rs2904130	G	176	0.349	174	0.379	184	0.351	0.88	0.4	0.6	0.88	0.4	0.4	1.00	1.0	0.8
rs34238957	0	169	0.385	172	0.390	180	0.375	0.98	0.9	0.1	0.94	0.7	0.1	0.96	0.8	0.8
rs34844435	0	175	0.963	174	0.934	184	0.973	1.83	0.08	0.2	2.53	0.01	0.04	1.38	0.4	0.5
rs34844435	1	175	0.037	174	0.066	184	0.022	0.55	0.08	0.2	0.31	0.003	0.01	0.58	0.2	0.2
rs35186448	2	175	0.191	174	0.239	181	0.199	0.76	0.1	0.5	0.79	0.2	0.9	1.05	0.8	0.6
rs3743073	G	175	0.386	175	0.383	177	0.376	1.01	0.9	0.08	0.97	0.8	0.08	0.96	0.8	1.0
rs3743074	G	175	0.386	175	0.383	184	0.372	1.01	0.9	0.08	0.96	0.8	0.1	0.94	0.7	0.9
rs3743075	T	176	0.384	175	0.383	184	0.372	1.00	1.0	0.08	0.96	0.8	0.1	0.95	0.8	0.9
rs3743077	T	176	0.426	170	0.465	176	0.395	0.86	0.3	0.5	0.75	0.06	0.9	0.88	0.4	0.6
rs3743078	C	176	0.190	173	0.237	179	0.201	0.76	0.1	0.5	0.81	0.2	1.0	1.07	0.7	0.6
rs3841324	del	152	0.424	163	0.457	177	0.387	0.88	0.4	0.5	0.75	0.06	1.0	0.86	0.3	0.5
rs41280050	A	175	0.020	175	0.011	180	0.019	1.77	0.4	0.3	1.72	0.4	0.4	0.97	1.0	1.0
rs472054	A	174	0.388	175	0.386	184	0.372	1.01	1.0	0.09	0.94	0.7	0.1	0.94	0.7	0.9
rs4887069	G	175	0.194	174	0.241	181	0.204	0.76	0.1	0.5	0.81	0.2	0.9	1.07	0.7	0.6
rs503464	A	164	0.195	171	0.240	176	0.207	0.77	0.2	0.5	0.83	0.3	1.0	1.08	0.7	0.6

rs555018	G	174	0.422	170	0.462	175	0.391	0.85	0.3	0.5	0.75	0.06	1.0	0.88	0.4	0.5
rs55781567	G	165	0.394	171	0.307	181	0.403	1.47	0.02	0.5	1.53	0.008	0.6	1.04	0.8	1.0
rs55783657	A	176	0.009	173	0.006	183	0.025	1.48	0.7	0.9	4.34	0.03	0.1	2.93	0.09	0.09
rs55787222[1]	4	162	0.506	166	0.551	164	0.503	0.83	0.2	0.2	0.82	0.2	0.09	0.99	0.9	0.8
rs55787222	5	162	0.071	166	0.087	164	0.052	0.80	0.4	0.8	0.57	0.07	0.2	0.72	0.3	0.3
rs55787222	2	162	0.423	166	0.361	164	0.436	1.29	0.1	0.09	1.37	0.05	0.1	1.06	0.7	0.8
rs55853698	G	168	0.390	171	0.304	180	0.400	1.46	0.02	0.9	1.53	0.008	0.6	1.04	0.8	0.6
rs55919125	T	176	0.043	174	0.032	183	0.055	1.36	0.4	0.3	1.77	0.1	0.05	1.30	0.5	0.4
rs55952530	A	170	0.009	171	0.020	178	0.017	0.43	0.2	0.3	0.82	0.7	1.0	1.93	0.3	0.3
rs55958820	T	176	0.017	172	0.012	180	0.017	1.47	0.5	0.8	1.44	0.6	0.9	0.98	1.0	0.9
rs56182392	A	158	0.013	168	0.012	173	0.014	1.06	0.9	0.8	1.22	0.8	0.6	1.14	0.8	0.8
rs564585	G	175	0.223	173	0.301	183	0.221	0.67	0.02	0.2	0.66	0.02	0.2	0.99	1.0	0.8
rs569207	T	176	0.190	175	0.237	182	0.203	0.76	0.1	0.5	0.82	0.3	1.0	1.09	0.7	0.5
rs578776	T	175	0.223	175	0.300	179	0.209	0.67	0.02	0.2	0.62	0.006	0.09	0.92	0.7	0.8
rs60706203	del	165	0.409	154	0.399	161	0.379	1.04	0.8	0.09	0.92	0.6	0.3	0.88	0.4	0.5
rs615470	T	176	0.386	174	0.388	184	0.372	0.99	1.0	0.09	0.94	0.7	0.1	0.94	0.7	0.8
rs621849	G	176	0.426	175	0.466	168	0.405	0.85	0.3	0.5	0.78	0.1	0.7	0.92	0.6	0.8
rs647041	T	172	0.427	172	0.471	177	0.395	0.84	0.3	0.6	0.74	0.04	0.7	0.88	0.4	0.4
rs6495306	G	176	0.426	175	0.466	168	0.405	0.85	0.3	0.5	0.78	0.1	0.7	0.92	0.6	0.8
rs660652	A	169	0.385	171	0.389	179	0.374	0.98	0.9	0.1	0.94	0.7	0.1	0.96	0.8	0.9
rs680244	A	176	0.426	156	0.474	168	0.405	0.82	0.2	0.7	0.75	0.07	0.9	0.92	0.6	0.8
rs684513	G	172	0.180	156	0.215	168	0.188	0.80	0.3	0.7	0.84	0.4	0.8	1.05	0.8	0.6
rs7178270	C	174	0.402	172	0.445	177	0.370	0.84	0.3	0.8	0.73	0.04	0.6	0.87	0.4	0.4
rs8040868	C	175	0.429	174	0.382	182	0.431	1.21	0.2	0.04	1.23	0.2	0.02	1.01	0.9	0.8
rs8192475	T	176	0.040	175	0.083	184	0.027	0.46	0.02	0.05	0.31	0.001	0.004	0.67	0.3	0.4
rs8192479	T	176	0.028	175	0.026	182	0.038	1.11	0.8	0.7	1.52	0.3	0.8	1.37	0.5	0.5
rs8192482	T	175	0.380	173	0.292	183	0.396	1.49	0.01	1.0	1.59	0.003	1.0	1.07	0.7	1.0
ss107794645	C	172	0.451	161	0.332	183	0.432	1.65	0.002	0.1	1.53	0.007	0.5	0.93	0.6	0.4

Padj - p-value after adjustment for the effect of rs1051730

1-Allele number refers to number of copies of 4bp repeat.

Table 3.2 - Equivalence classes for SNPs with minor allele frequency greater than 5%

Class	A		B		C		D		E		F	
Head	rs1051730		rs680244		rs1948		rs8192475		rs578776		rs569207	
	rs16969968	1.00	rs34238957	0.82	rs2904130	0.92	rs34844435	0.88	rs564585	0.99	rs35186448	0.99
	rs8192482	1.00	rs3841324	0.91			rs12898919	1.00			rs503464	0.86
	rs55853698	0.93	rs60706203	0.87			rs12899226	1.00			rs3743078	0.99
	rs55781567	0.93	rs621849	1.00			rs12907519	0.93			rs7170068	0.87
			rs6495306	1.00			rs12911814	1.00			rs684513	0.83
			rs555018	1.00							rs28534575	0.83
	rs55787222	0.64	rs647041	0.99							rs4887069	0.96
	rs8040868	0.79	rs615470	0.82							rs13329271	0.90
	ss107794645	0.69	rs3743075	0.81								
			rs3743074	0.81								
			rs3743073	0.81								
			rs3743077	1.00								
			rs660652	0.82								
			rs472054	0.82								
			rs7178270	0.80								

All variants with frequency greater than 5% were grouped into equivalence classes based on r2>0.8.

A lead SNP for each class was chosen, and r^2 for each variant to that SNP is listed.

Three variants do not fit into these classes. They are listed separately under class A, to which each has the strongest LD.

Table 3 - Association Results for Equivalences Classes in Larger Samples

			LQS		ND		LC		ND against LQS			LC against LQS			LC against ND		
	Class	Allele	N	freq	N	freq	N	freq	OR	Padj	Padj 1	OR	Padj	Padj 1	OR	Padj	Padj 1
rs1051730	A	T	4676	0.309	1950	0.384	669	0.404	1.40	$7.1x10^{-15}$	-	1.52	$1.5x10^{-11}$	-	1.09	0.2	-
rs680244	B	A	4680	0.444	1950	0.405	669	0.407	0.85	$9.2x10^{-5}$	0.3	0.86	0.01	0.03	1.01	0.9	0.2
rs1948	C	T	4674	0.350	1950	0.318	669	0.333	0.87	0.001	0.4	0.93	0.2	0.006	1.07	0.3	0.04
rs8192475	D	T	4674	0.053	1948	0.046	668	0.042	0.87	0.2	0.9	0.78	0.09	0.5	0.90	0.5	0.6
rs569207[1]	F	T	4675	0.249	1950	0.213	669	0.191	0.82	$3.0x10^{-5}$	0.2	0.71	$3.1x10^{-6}$	0.03	0.87	0.09	0.2

1. rs569207 allele T is tagged by a haplotype of allele C at rs1051730 and allele G at rs680244 (r^2=0.98)

P - includes correction for relatedness among groups.

Padj 1 adjusts the p-value for the significant effect of rs1051730

Table 4 - Association Results for Insertion/Deletions and Microsatellites

		LQS		ND		LC		ND against LQS			LC against LQS			LC against ND		
	Allele	N	freq	N	freq	N	freq	OR	P	Padj	OR	P	Padj	OR	P	Padj
rs1051730	T	608	0.306	1623	0.359	567	0.384	1.27	0.001	0.001	1.42	$8.5x10^{-5}$	$8.5x10^{-5}$	0.98	0.8	0.4
rs3841324	del	608	0.433	1623	0.403	567	0.399	0.88	0.08	0.8	0.87	0.1	0.4	0.95	0.5	0.8
rs60706203	ins	608	0.398	1623	0.385	567	0.374	0.95	0.4	0.2	0.90	0.2	0.2	1.15	0.05	0.2
rs55787222	2	608	0.382	1623	0.404	567	0.438	1.10	0.2	0.004	1.26	0.006	0.2	0.89	0.1	0.4
rs55787222	4	608	0.536	1623	0.514	567	0.484	0.91	0.2	0.1	0.81	0.01	0.6	0.94	0.7	0.9
rs55787222	5	608	0.076	1623	0.075	567	0.071	1.00	1.0	0.5	0.94	0.7	0.7	1.05	0.9	0.8
rs55787222	7	608	0.004	1623	0.006	567	0.006	1.43	0.5	0.4	1.50	0.5	0.4	1.11	0.1	0.1

The results for the T allele of rs1051730 within the sample genotyped for length variants is included here in the table for comparison.

P - includes correction for relatedness among groups.

Padj - P-value after adjustment for the effect of rs1051730.

Table 3.5 - Association analysis of CHRNA5 expression with variants in the LD block.

			Whole Blood						Subcutaneous Adipose					
marker	allele	class	r	LCL	UCL	P	$adjP_1$	$adjP_2$	r	LCL	UCL	P	$adjP_1$	$adjP_2$
rs3841324	0	B	0.72	0.67	0.76	$3.7x10^{-71}$	-	-	0.73	0.68	0.77	$1.9x10^{-63}$	-	-
rs1051730	4	A	-0.54	-0.60	-0.47	$2.0x10^{-34}$	0.006	-	-0.53	-0.60	-0.45	$3.9x10^{-28}$	0.02	-
rs680244	1	B	0.71	0.66	0.75	$4.3x10^{-69}$	0.03	0.1	0.71	0.65	0.75	$3.6x10^{-58}$	0.07	0.3
rs1948	4	C	0.58	0.51	0.63	$1.3x10^{-40}$	0.8	0.5	0.56	0.49	0.63	$2.0x10^{-32}$	0.8	0.6
rs8192475	4	D	0.26	0.17	0.34	$3.0x10^{-8}$	0.03	0.04	0.23	0.13	0.33	$5.3x10^{-6}$	0.5	0.6
rs578776	4	E	-0.14	-0.23	-0.05	0.002	0.06	0.5	-0.10	-0.20	0.00	0.004	0.08	0.7
rs569207[1]	4	F	-0.23	-0.31	-0.14	$1.3x10^{-6}$	0.06	0.3	-0.19	-0.29	-0.09	$1.6x10^{-4}$	0.09	0.5

1. rs569207 allele T is tagged by a haplotype using allele G at rs680244 and allele T at rs578776 (r2=0.99)

$adjP_1$ - adjusted for the effect of rs3841324; adjP2 - adjusted for effects of both rs3841324 and rs1051730

Association analysis of the effect of rs3841324 genotype on expression of CHRNA5. Other variants within the block were tested as well by including the head of each

equivalence class.

Table 3.6 - Demographics: Sequencing Cohort

	N	Sex (M/F)	Age (yrs)
Low Quantity Smokers	175	57/118	55.8±18.4
Nicotine Dependence	176	79/97	50.6±10.4
Lung Cancer	184	98/86	72.6±10.8

Table 3.7 - Descriptive information on all variants from sequencing

CHRNA5

refSNP ID	Position (b36)	Major Allele	Minor Allele	Minor Allele Freq	Function	aa change
ss107794609	76644592	G	T	0.2%	near-gene_5	
rs3841324	76644866	22bp[1]	-	42.2%	near-gene_5	
rs56182392	76644932	G	A	1.3%	near-gene_5	
rs503464	76644949	T	A	21.4%	near-gene_5	
rs55853698	76644992	T	G	36.5%	utr-5	
rs55781567	76645039	C	G	36.8%	utr-5	
ss107794620	76645329	G	A	0.2%	intron	
rs684513	76645453	C	G	19.4%	intron	
rs6495306	76652946	A	G	43.3%	intron	
rs680244	76658341	G	A	43.4%	intron	
rs621849	76659914	A	G	43.3%	intron	
ss107794606	76660068	A	C	0.6%	intron	
ss107794638	76660152	G	A	0.2%	intron	
rs569207	76660172	C	T	21.0%	intron	
ss107794639	76660615	A	G	0.6%	intron	
rs55982512	76666111	C	T	0.4%	intron	
rs555018	76666295	A	G	42.5%	intron	
rs647041	76667534	C	T	43.1%	intron	
ss107794648	76667613	TC	-	0.1%	intron	
rs12898919	76667630	G	C	4.8%	intron	
rs2229961	76667805	G	A	1.1%	non-synon	V->I
rs56201623	76669057	C	T	0.1%	intron	
ss107794603	76669153	T	C	0.4%	intron	
ss107794604	76669479	C	T	0.3%	synon	
rs16969968	76669978	G	A	36.0%	non-synon	D->N
ss107794605	76670139	C	T	0.4%	intron	
ss107794607	76672422	G	C	1.0%	intron	
ss107794649	76672960	ACT	-	0.1%	utr-3	
rs615470	76673041	C	T	38.2%	utr-3	
rs8192483	76673202	G	A	0.1%	utr-3	
rs55783657	76673211	G	A	1.3%	utr-3	
rs8192482	76673251	C	T	35.7%	utr-3	
rs564585	76673280	A	G	24.8%	utr-3	
ss107794608	76673349	G	A	0.1%	utr-3	

CHRNA3

refSNP ID	Position (b36)	Major Allele	Minor Allele	Minor Allele Freq	Function	aa change
rs12899226	76674491	A	C	4.9%	near-gene_3	
rs55736590	76674548	C	T	0.7%	near-gene_3	
rs34238957	76674769	-	CTCT	38.3%	utr-3	
rs660652	76674885	C	T	38.2%	utr-3	
ss107794646	76675046	T	C	0.2%	utr-3	
rs472054	76675047	C	T	38.2%	utr-3	
rs35186448	76675292	-	CCCC	20.9%	utr-3	
rs56113144	76675404	C	T	0.3%	utr-3	
rs578776	76675453	G	A	24.4%	utr-3	
ss107794615	76676126	T	A	0.2%	non-synon	I->N
rs56403513	76680840	C	T	0.1%	synon	
ss107794613	76681059	C	T	0.1%	synon	
ss107794633	76681388	C	T	0.1%	non-synon	H->Y
rs1051730	76681392	G	A	35.9%	synon	
rs55958820	76681410	C	A	1.5%	synon	
rs8192480	76681473	T	C	0.1%	synon	
ss107794647	76681494	20bp[2]	-	0.1%	frameshift	
ss107794614	76681537	C	T	0.1%	non-synon	P->L
rs3743078	76681812	C	G	20.9%	intron	
rs3743077	76681949	G	A	42.8%	intron	
ss107794634	76696117	G	A	0.9%	intron	
rs4887069	76696123	C	C	21.3%	intron	
rs8192479	76696451	G	A	3.1%	synon	
rs3743075	76696505	G	A	37.9%	synon	
rs3743074	76696533	T	C	38.0%	intron	
rs3743073	76696592	A	C	38.1%	intron	
rs41280050	76696610	C	T	1.7%	intron	
ss107794612	76696706	G	A	0.8%	intron	
ss107794635	76696791	G	C	0.4%	intron	
ss107794650	76698092	-	A	4.1%	intron	
ss107794650	76698092	A	-	0.2%	intron	
rs8040868	76698234	A	G	41.4%	synon	
rs8192475	76698283	G	A	5.0%	non-synon	R->H
ss107794610	76698482	C	T	0.1%	intron	
ss107794611	76698486	G	C	0.2%	intron	
rs7170068	76699996	C	T	24.3%	intron	
ss107794616	76700023	G	A	0.1%	intron	
ss107794617	76700031	A	G	0.1%	intron	
rs12907519	76700097	A	G	5.0%	intron	
rs60706203	76700140	AGC[3]	-	39.6%		
ss107794636	76700422	C	A	0.2%	near-gene_5	
rs55787222	76700426	(CGCC)2-7[4]			near-gene_5	
ss107794621	76700489	C	G	0.1%	near-gene_5	
ss107794622	76700598	C	T	0.8%	near-gene_5	
ss107794623	76700882	T	C	0.1%	near-gene_5	
ss107794637	76700886	T	C	0.1%	near-gene_5	
ss107794624	76700991	A	G	0.2%	near-gene_5	
rs12911814	76701037	T	G	5.1%	near-gene_5	
rs13329271	76701283	T	G	10.0%	near-gene_5	

CHRNB4

refSNP ID	Position (b36)	Major Allele	Minor Allele	Minor Allele Freq	Function	aa change
rs2904130	76703675	C	G	36.0%	near-gene_3	
ss107794618	76704149	C	G	0.8%	utr-3	
rs55952530	76704369	G	A	1.5%	utr-3	
rs1948	76704452	C	T	34.6%	utr-3	
ss107794644	76704905	C	T	0.1%	intron	
rs7178270	76708130	G	C	40.5%	intron	
rs56317523	76708396	C	T	0.4%	non-synon	A->V
rs56235003	76708655	C	T	0.8%	non-synon	R->C
rs3743072	76708815	C	T	0.1%	synon	
rs55919125	76709247	C	T	4.3%	synon	
rs56218866	76709282	A	G	0.1%	non-synon	S->G
rs56095004	76709293	G	A	0.7%	non-synon	R->Q
ss107794619	76709462	C	T	0.1%	intron	
ss107794640	76710240	A	G	0.3%	intron	
ss107794641	76710248	A	G	0.2%	intron	
rs12914008	76710558	C	T	3.5%	non-synon	T->I
ss107794625	76710749	A	G	0.1%	intron	
rs28534575	76710898	A	C	21.1%	intron	
ss107794642	76710923	G	A	0.1%	intron	
rs12440298	76714642	A	C	0.2%	intron	
ss107794630	76714647	T	G	0.2%	intron	
ss107794629	76714668	A	G	0.1%	intron	
ss107794628	76714715	C	T	0.1%	intron	
ss107794627	76714923	C	T	0.3%	non-synon	R->S
ss107794626	76715161	C	T	0.1%	intron	
ss107794643	76720729	G	C	0.1%	near-gene_5	
ss107794632	76720778	C	T	0.2%	near-gene_5	
ss107794631	76721201	A	T	0.2%	near-gene_5	
ss107794645	76721371	G	C	40.7%	near-gene_5	

[1] CTATTTCCCTCTGGCCCCGCCC

[2] ATCGATTTTCGCCTTATCGT

[3] The major allele contains 7 copies of AGC, the minor 6. With genotyping of 2798 individuals, one individual was identified with 8 copies.

[4] Frequencies for alleles of rs55787222 are based on specific genotyping of this variant in 2935 individuals.
Frequencies are as follows - 2: 41.6%, 3: 0.07%, 4: 50.4%, 5: 7.4%, 6: 0.02%, 7: 0.6%

Table 3.8 - All variants with frequency less than 1%.

		Number of Carriers of Minor Allele		
ref id	Allele	LQS	ND	LC
rs55665143	(-A)	0	0	2
ss107794647	(-20bp)1	0	1	0
ss107794648	(-TC)	1	0	0
ss107794649	(-ACT)	0	0	1
rs55982512	T	1	1	2
ss107794603	C	1	2	1
ss107794604	T	0	0	3
ss107794605	T	4	0	0
ss107794606	C	1	2	3
ss107794607	C	4	1	5
rs8192483	A	0	1	0
ss107794608	A	1	0	0
ss107794609	T	1	1	0
ss107794610	A	1	0	0
ss107794611	G	1	0	1
ss107794612	T	2	4	3
rs56403513	A	0	0	1
ss107794613	A	1	0	0
rs8192480	G	0	0	1
rs55736590	A	3	1	3
rs56113144	A	0	2	1
ss107794615	T	1	0	1
ss107794616	T	0	1	0
ss107794617	C	1	0	0
ss107794618	G	2	2	4
ss107794619	T	0	0	1
rs56218866	G	0	0	1
rs56095004	A	0	2	5
rs3743072	T	0	0	1
rs56235003	T	4	1	3
rs56317523	T	0	1	3
ss107794620	A	1	0	1
ss107794621	C	1	0	0
ss107794622	A	3	1	2
ss107794623	G	0	0	1
ss107794624	C	1	0	0

ss107794625	G	0	0	1
ss107794626	T	1	0	0
ss107794627	T	0	2	1
ss107794628	T	0	1	0
ss107794629	G	0	1	0
ss107794630	G	0	0	2
rs12440298	C	0	1	1
ss107794631	T	0	2	0
ss107794632	T	1	0	1
ss107794646	C	2	0	0
ss107794633	A	0	0	1
ss107794634	A	6	1	2
ss107794635	G	2	1	1
ss107794636	T	1	0	0
ss107794637	G	0	1	0
ss107794638	A	2	0	0
ss107794639	G	3	2	1
rs56201623	T	0	0	1
ss107794640	G	1	1	0
ss107794641	C	0	2	0
ss107794642	A	0	1	0
ss107794643	C	1	0	0
ss107794644	T	0	1	0

[1]
ATCGATTTTCGCCTTATCGT

Supplemental Table 3 - Demographics for cohorts used in analyses including additional genotyping

A - Length Polymorphisms	N	Male/Female	Age (years)
Nicotine Dependence	1623	602/1021	50.4±11.2
Lung Cancer	567	291/276	70.6±11.0
Low Quantity Smokers	608	192/416	58.4±18.2
B - rs578776			
Nicotine Dependence	2161	758/1403	50.1±11.3
Low Quantity Smokers	865	283/582	57.7±18.8
C - Illumina			
Nicotine Dependence	1950	689/1261	51.0±11.0
Lung Cancer	669	340/329	70.6±11.0
Low Quantity Smokers	4681	1203/3478	63.9±19.1

Age is in years ± S.D.

Supplemental Table 4 - Full list of sequencing primers used

	Primer Sequence
CHA5.01a.F	TCTGGAGGGATTTGAGACG
CHA5.01a.R	GAGTTGTGATCAGAAAGAAACAAGC
CHA5.01b.F	TCAAGCTGCGACTGGTACTT
CHA5.01b.R	TTTGGCCCAGTCCGGTCTAT
CHA5.01c.F	GTCTTTGCCTTCCTGGAACT
CHA5.01c.R	GCGACCAGCTGGACCAAG
CHA5.01d.F	CCCAAGAGTTCGCGTTCC
CHA5.01d.R	TGCTGGACAAGAGTGGCATC
CHA5.02a.F	GAGCTCAGCATATTCCAATAGTCA
CHA5.02a.R	GAAGGATATGCACACCTACCAC
CHA5.02b.F	CATGGCCCAGGTTGGAGT
CHA5.02b.R	CACAATCACTACTTTTCCCTGGT
CHA5.02c.F	CAGGATTATCTGAACCTTCTTCTATTG
CHA5.02c.R	CTTTTCCCTCCTTTCAAGCTA
CHA5.03a.F	GCTAAGCACATAAAATAACACTATGC
CHA5.03a.R	ACTAATTGGGTGCCCGTTTG
CHA5.03b.F	TGATGAATGAAATGTGGGTACTA
CHA5.03b.R	CCAGCCTTGGGTATGTCC
CHA5.04a.F	GCCAAATTGGGTGACAATAC
CHA5.04a.R	CAAAACGATGTCTGGTGTCC
CHA5.04b.F	TTCTTTGTTTTAAAGGAATGGATAG
CHA5.04b.R	TCTAGAAGTCTGCCAAATATGTAGTC
CHA5.05a.F	GCCTGAGTCTATTCTGTGTGTAAGG
CHA5.05a.R	CCTGTGATCCATCATAAGTCCAA
CHA5.05b.F	GTTTTGAAGGGACCAGTACG
CHA5.05b.R	GGTATGACTTTTGAAGATGATGGTATG
CHA5.05c.F	CGCCTGCCTCTCTTTTATACC
CHA5.05c.R	ATCACAGACCTCACGGACATC
CHA5.05d.F	CCGCAAGATATTTCTTCACACG
CHA5.05d.R	CAAGAAATTAAAACAATTCATCAGGTC
CHA5.06a.F	AAGTAAACACTACTGGGCAAGAA
CHA5.06a.R	CCTTGGGAGGCTTCACTTATT
CHA5.06b.F	ATAGCCCAGGTTCTTGATCG
CHA5.06b.R	CAGGAAAGTTTACATACCTTTACCAA
CHA5.06c.F	TGAACAGTTGGCTGTATGACTG
CHA5.06c.R	CACCACCCTGGCTAATTTCA
CHA5.06d.F	TGCATTTGGTAAAGGTATGTAAAC
CHA5.06d.R	GCCTTAAAGGAGTTCCCAAT
CHA5.06e.F	GCAACAAGAGCAAAACTCAGTC
CHA5.06e.R	GGGGTTCATGTAATTGTAGTGG
CHA3.01a.F	TCCCGCTTTCTCTCTTGTC
CHA3.01a.R	GCGGATCACGAGGTCAGA
CHA3.01b.F	CGGCTCACTGCAGGCTCT
CHA3.01b.R	GGCTCCAGGTCCCAGTCC
CHA3.01c.F	AACCTGGGACAGAAACTGAGTC
CHA3.01c.R	TACCTGGCAGCAGAGACAGC
CHA3.01d.F	CTCCAGGTCTGGGGTCTG
CHA3.01d.R	CAGTTTGGGAGCCAGTGC
CHA3.02a.F	CGTTTATAATCCTTATTTTGACTGGTA
CHA3.02a.R	TGAGACATGGACACCTCGAA
CHA3.02b.F	TGGAGGGGATGCTGTAGAT
CHA3.02b.R	ATGAAGCCTGGTCTTTAGGC
CHA3.02c.F	AGTGCTGGAAGCCACGAG
CHA3.02c.R	GGCTTCAATCTCAGGTTCGT
CHA3.03a.F	AGGAGTCTCCACCTGGCACTA
CHA3.03a.R	CCAGCACCTTACTTGTTATACAGC

CHA3.03b.F	CCAGGCTGATTCTTTTACCG
CHA3.03b.R	AGTCCAGCCAATGAGGTCAC
CHA3.04a.F	TGGCTCCAGGAGGATGAC
CHA3.04a.R	TTCGCCTTATCGTAGGACCA
CHA3.04b.F	AGGTGGACGACAAGACCAA
CHA3.04b.R	TCATGGTGAACAGGAGGTACT
CHA3.04c.F	ACCCTGTGCATTTCTGTCC
CHA3.04c.R	TGACACTTTGGATGGCTTC
CHA3.04d.F	AGGACGGGATGTGTGGTTACT
CHA3.04d.R	CATCCTAATCTAGTCTATTGGCATC
CHA3.05a.F	ATCCTGCCATGGAAGCCTA
CHA3.05a.R	AAACAAAGCTGGTAGCTTGATAAC
CHA3.05b.F	TGCAGATTCAAGATGATTGG
CHA3.05b.R	CCTTCAAAGAGATTATGGGCTA
CHA3.05c.F	TGTCAAAATGTGATTCTATGTGATTAG
CHA3.05c.R	CAGAGATAGCTAGTCCCTTCAGTC
CHA3.05d.F	CATTTCAGAGAGCTTCAACTACTTC
CHA3.05d.R	TGAGTTGAGGCTAGGGTACTGC
CHA3.05e.F	CAATCACGCTGGGAATAGGTT
CHA3.05e.R	CATCCCTGATCCCTCTTATACTAC
CHA3.05f.F	CACAGAGACTGTTTGAATCTTGTC
CHA3.05f.R	TTCAGGTGTGACCAGGTAGC
CHB4.01a.F	CATGTCAGATTATAGCAAAATGACG
CHB4.01a.R	CATGGGCCTTCCAATCTG
CHB4.01b.F	ACCCGCAAGGGAATGGTACT
CHB4.01b.R	GGGCACCCTTGGTACAGC
CHB4.01c.F	GCGAATGAACCTGAGATGAC
CHB4.01c.R	GACCAGGAAGAAAAGGACCA
CHB4.01d.F	GACCGGCGCTCACTCGAC
CHB4.01d.R	ATCTGGCCGGGACAATCT
CHB4.02a.F	TGAACCAAGGGAGAGGTCAA
CHB4.02a.R	TCTGCACCTACCACGCTGA
CHB4.02b.F	GAGGAAAAGCTGATGGACGA
CHB4.02b.R	AGAGAATAGGGTGGGGCTGTA
CHB4.03a.F	GGATCAGAGAGCAAAGTGTCA
CHB4.03a.R	GTAGCGGGAGCTGTTCCA
CHB4.03b.F	TGTCTGGCTGAAACAGGTAAG
CHB4.03b.R	GCAACAGAGGCTCAAAAGGATAG
CHB4.04a.F	CTACTTGGGAGGCCGAGA
CHB4.04a.R	AGGACCATGTCTATCTCCGTGT
CHB4.04b.F	GGGGTGCTTTGATGTTAGG
CHB4.04b.R	AGGGCCACTATGTCCCACTC
CHB4.04c.F	TTGATAGTCCGGTCCAACG
CHB4.04c.R	TGGTGAACATGAGGTACTTGC
CHB4.04d.F	CTCGTCTTCTACCTGCCATC
CHB4.04d.R	CTCCTGCACATCCTGTCG
CHB4.04e.F	CCTCTGCAGCTTCCAAGTC
CHB4.04e.R	GGGTACCCACGGCAGTATC
CHB4.05a.F	ACAAGCCCTCACATTCCTAGC
CHB4.05a.R	CCAATGCTCACATATTTACTTAGGG
CHB4.05b.F	TGGAAGTACGTGGCTATGGT
CHB4.05b.R	ATATGGCAAATGCCAAGC
CHB4.05c.F	TTCGTAGCAGCACCTACTATGC
CHB4.05c.R	CAGGGTAACGTGACTGTAGGG
CHB4.05d.F	TATATCGCCCAGGCTCAA
CHB4.05d.R	ACGAATGTGAAGGAGCAGGT
CHB4.05e.F	TATGCCTGGCCTTCCCTAT
CHB4.05e.R	GAGAGAGAAGTGGAAAGTGACC

Supplemental Table 5 - Build 36 positions for regions sequenced

Gene	Region	Build 36 position
CHRNA5	5' Flanking & Exon 1	76643986-76645528
	Exon 2	76659873-76660680
	Exon 3	76665714-76666400
	Exon 4	76667349-76668117
	Exon 5	76668894-76670363
	Exon 6 & 3' Flanking	76672141-76673771
CHRNB4	5' Flanking & Exon 1	76720345-76721584
	Exon 2	76714503-76715286
	Exon 3	76710160-76711022
	Exon 4	76708007-76709537
	Exon 5 & 3' Flanking	76703378-76704963
CHRNA3	5' Flanking & Exon 1	76699749-76701312
	Exon 2 &3	76697716-76698675
	Exon 4	76696032-76696844
	Exon 5	76680349-76682010
	Exon 6 & 3' Flanking	76674343-76676459

Chapter 4

Discussion

A. Summary of Results

Analysis of genetic association of the LD block containing rs1051730 with dependence to several substances demonstrated no effect of the smoking risk variant on dependence to alcohol, amphetamines, cannabis, opiates or sedatives/hypnotics. The previously reported association for alcohol dependence with other SNPs within the LD block, such as rs680244 was thus not replicated(77). Given that the sample available for cocaine dependence was too small to report on, we could not directly address with strong inference the findings of Grucza et al.(78) with regard to cocaine dependence. We also had smoking quantity information available for a substantial proportion of our alcohol dependence cohort. Within this group we saw significant association of the smoking risk variant rs1051730 to smoking quantity within the alcohol dependence cohort, and the effect size is the same as that seen in the general population.

Characterization of the entire nAchR gene cluster was performed in a sample of lung cancer patients, nicotine dependent smokers, and low quantity smokers. No variants were identified with stronger association to either phenotype than rs1051730. The non-synonymous variant rs16969968 was found to be equivalent to rs1051730 in Iceland. Linkage disequilibrium between SNPs in the block was characterized and defined.

Analysis of gene expression in an independent cohort indicate that the promoter insertion/deletion rs3841324 is strongly associated with expression of CHRNA5 in humans. Genotyping of this variant within the lung cancer, nicotine dependence and low quantity smoking groups indicates that there is no risk of this variant for either phenotype. Interestingly, we found that the risk allele of rs1051730/rs16969968 appears only on the haplotype background which contains the low expression variant of rs3841324, which would indicate that CHRNA5 mRNA containing the amino acid substitution caused by

rs16969968 will be transcribed less frequently, although the mechanism for that regulation is not known. Regardless of whether rs3841324 genotype influences the risk for disease carried by rs1051730, the transcriptional regulation mediated by rs3841324 may influence studies which attempt to characterize the functional effects of the variant.

B. Other substances of abuse

We do not see evidence of association of any variants within this gene cluster with dependence to substances other than nicotine. In the case of alcohol dependence our sample is large enough to ensure that power is adequate to detect association. However, for other substances our samples are considerably smaller, and there is less statistical power. Small sample sizes in individual studies is a frequent limitation of studies within this field. All of the data is provided such that it can be used in the future for meta analyses.

We do see association of rs1051730 with smoking quantity among alcoholics. For other dependence diagnoses the sample sizes become too small for the statistical model used in the alcohol dependence cohort and the general population to be an appropriate fit. The effect size estimate from the regression analysis is the same when the model is applied to alcoholics as it is when applied to the rest of the population included as controls in this study. This is actually quite intriguing. At first glance it appears that the variant has no relationship to alcohol dependence and exactly the same effect on smoking in these smokers as in controls. In some sense this is indeed true. However, considering the relationship between alcohol dependence and smoking quantity reveals a more complex problem which needs to be addressed further. Smoking quantity is higher among alcoholics than the general population. Thus the distribution across the categories used

for this analysis is quite different in alcoholics compared to the general population. This means that in order for the overall frequency of the variant in alcoholics to be the same as in the general population, the frequency of the variant for any given smoking quantity level must be lower, as indeed it is. Had there been a significant association of rs1051730 to alcohol dependence, we would be faced with a complex situation similar to that involving lung cancer and peripheral arterial disease. Do we see significant association of the variant with the disease? Or do we simply have a case group which is enriched for individuals whose smoking behavior leads them to have a high frequency of the variant. The combination of results we see, where there is no association of the variant with alcohol dependence, but there is association of the variant with smoking quantity among alcoholics, would seem to indicate that the variant affects smoking in a way that is independent from the effect which increased alcohol consumption or alcohol dependence has on smoking. However, interpretation of these results should be done with care. A better understanding of the phenotypic differences in smoking behavior and nicotine dependence between alcoholics (or other drug abusers) and non-abusers in concert with studies of this variant might help elucidate this complex interplay

C. Sequencing and Expression Analysis

The sequencing of all three genes within the cluster provides the first report of a sequencing of genes in the cluster which is also aimed at lung cancer cases. Lung cancer cases are primarily included as a group of subjects with very significant lifetime tobacco consumption. However, given the continuing uncertainty over whether the effect seen in lung cancer is due only to an increase in smoking exposure, or to an increased susceptibility to *nicotine-induced* carcinogenesis, the inclusion of sequencing data from

lung cancer cases allows some examination of potential differences in frequencies of variants between lung cancer and nicotine dependence. We do not see any significant differences. However, the power of these samples, which are relatively large for sequencing studies (approximately 180 cases per group), is limited for detecting association to rare variants. Our study does not attempt to detect association with rare variants, but rather to identify and describe them.

Analysis of the region has turned out to be quite complex, and future studies which characterize regions with known associations may find similar challenges. Because we have larger cohorts available for the phenotypes tested, we were able to follow up on potentially interesting variants seen in sequencing data or analysis. One variant showed significant association in the sequencing sample, but analysis of data for a tagging SNP genotyped in the larger sample demonstrated that this effect does not replicate.

Because there is confirmed association with rs1051730 and many SNPs are correlated with that SNP, each association was adjusted for the effect of rs1051730. This led to some results which were difficult to interpret. Even from the first set of tests, within the sequencing sample, this study faced a large multiple testing penalty. Statistical testing began with 50 variants and 3 phenotype tests. The presence of 6 equivalence classes which account for most of the observed variants reduces the effective number of tests substantially. The study is primarily descriptive, and it is not beneficial here to simply state that associations do not survive multiple testing correction. The burden of multiple testing throughout the study does, however, makes it likely that most if not all findings of nominal significance are in truth noise. We are careful therefore not to overinterpret the results of association analyses where adjustment for the effect of rs1051730 leads to significant results with markers which showed no association prior to

adjustment. This occurs in 2 cases, rs1948 in lung cancer, and the 2-copy allele of the microsatellite rs55787222, in the promoter region of CHRNA3 in nicotine dependence.

The case is stronger for the microsatellite allele. Whether or not there is actually a true effect from this variant, the case is a useful exercise in understanding how such interactions might occur and how they present. The risk allele of rs1051730 is fixated on the background of the microsatellite allele containing 2 copies of the 4bp repeat unit. This allele occurs at 5-8% frequency without the risk allele of rs1051730. In isolation the strong LD with rs1051730 leads to some risk being observed with this allele. However, when the effect of rs1051730 is taken into account, it is then seen that this allele is quite protective in the absence of the risk allele of rs1051730. If we take this case as a model and do not concern ourselves with the absolute statistical significance of this test, this finding would lead to at least two possible conclusions: 1) This microsatellite allele in fact has a mildly protective effect on nicotine dependence on its own, but because the risk allele of rs1051730 has appeared only on this background, the substantially larger risk contributed by this variant masks the protective effect of the microsatellite allele (given the location of the microsatellite, in the promoter of CHRNA3, alteration of CHRNA3 expression would be a possible functional effect of the variant); 2) The haplotype which contains this microsatellite allele and the protective allele of rs1051730 is a tag for yet another variant, which is protective, but which is not captured elsewhere by our sequencing. The latter possibility is an unlikely, but certainly not impossible. In practice of course, a protective effect of this variant is dominated by the risk contributed by rs1051730. However, combined with the reports of very different allele frequencies for the variants rs1051730 and rs16969968 in populations of different ethnicities, there is the

possibility that the linkage disequilibrium structure in another population could shift the weight of such an effect.

The final contribution of this work is to firmly establish the effect of rs3841324, a CHRNA5 promoter insertion/deletion, on expression of CHRNA5 *in vivo*. The variant is strongly associated with expression of the gene in both whole blood and subcutaneous adipose tissue samples, with the short allele leading to higher expression. The effect is so strong that association can be picked up with many variants within the LD block due to their correlation with rs3841324. For variants in very close LD it would be difficult to determine whether one variant had a stronger effect than another. However, the functional study of Buckland et al.(99) on promoter variants provides compelling evidence that this insertion/deletion is indeed the variant regulating expression. The risk allele of rs1051730/rs16969968 appears only with the low expression allele of rs3841324. Analysis of rs3841324 genotypes in nicotine dependence and lung cancer demonstrates there is no risk carried by this variant for either phenotype. Even though the variant does not seem to affect risk for either disorder, its correlation with rs16969968 in conjunction with its strong effect on expression make it important to recognize when pursuing functional studies of rs16969968. The amino acid substitution caused by this variant will only occur in CHRNA5 transcripts which should be expressed at low levels resulting from their fixation on this haplotype background. More than 50% of transcripts from the low expression background will contain the missense mutation in Caucasian populations.

D. Discussion

A significant issue for future studies, highlighted in this study, is the question of how to analyze the role of rare variants. The era of genome-wide association studies has

identified many common variants that cause risk for common diseases, but much of the genetic risk for these diseases remains unaccounted for. The chips used for genome-wide association studies have primarily focused on common variants. Variants which occur in populations at or around 1% frequency present a challenge for both genotyping and analysis. Sequencing will be the gold standard for genotyping of rare SNPs, but costs and labor remain prohibitive. Next generation sequencing techniques provide dramatic improvements in the speed and flexibility of sequencing over traditional techniques, enabling the search for rare risk variants without specific *a priori* hypotheses possible, but still not easy. An approach being developed at Decode will combine the re-sequencing with a long range haplotype phasing bioinformatics tool under development(101). Long range phasing will allow the determination of carrier status for specific haplotypes harboring rare variants identified through sequencing. Then other carriers of the same long range haplotype in the population can be identified as most likely having the rare variant, without genotyping the entire population.

The question of what precisely is the effect of this variant, either biologically or phenotypically will not be answered by population genetics studies. Indeed, using population genetics we cannot with certainty determine which variant within the region is responsible for the observed associations. Additional studies with increasing sample size may identify additional variants in the region. The nonsynonymous variant rs16969968 is most likely to be the causal variant simply because it is most likely to be of functional significance. Working under that assumption the spotlight is cast on the α5 nAchR subunit. It is a challenging subunit to work with as described previously. It is expressed widely throughout not only brain but nearly all tissues in the body. Yet, it is never the

only α subunit in a nAchR. It is the only α subunit which has a purely modulatory role. It is expressed in the VTA, and thus receptors containing it could play an important role in the addiction pathway there. Differences between α4/β2 receptors containing α5 subunits and those that do not have been described(89) including increased calcium permeability, decreased concentration of agonist needed to desensitize, and marked increase in affinity. If the mutation were to affect any of these key features of nAchRs, the impact would be on multiple subsets of receptors. It is unlikely, given the relatively subtle phenotype resulting from the mutation and the location of the amino acid, that the mutation would affect core elements of the receptor such as the ligand binding site, or the residues lining the channel pore. However, the mutation may effect the kinetics of the ion channel through, for example, stabilizing one conformation of the receptor. This could lead to an increase in time spent in the open or the desensitized state. However, this is just one possible example. In combination with other subunits the α5 subunit may have a different role. Further research is needed to define whether this mutation does directly affect channel properties.

We have approached the study of this region with very large cohorts and detailed information, but it is difficult to determine a specific effect of the variant on one aspect of a phenotype as broad as nicotine dependence is. It is certainly highly unlikely that the only effect, or the true effect, of the variant is that for which it was identified, ie. increasing smoking quantity by one cigarette per day. Brain imaging studies and animal models will probably prove to be useful tools in eventually identifying the mechanism by which this variant increases an individual's smoking.

Chapter 5

References

1. Annual smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 1997-2001. MMWR Morb Mortal Wkly Rep. 2005 Jul 1;54(25):625-8.
2. WHO. Economics of tobacco control. Available at http://www.who.int/gb/fctc/PDF/wg1/e1t2.pdf.
3. Grant BF, Hasin DS, Chou SP, Stinson FS, Dawson DA. Nicotine dependence and psychiatric disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. Arch Gen Psychiatry. 2004 Nov;61(11):1107-15.
4. Cigarette Smoking Among Adults --- United States, 2006. MMWR Morb Mortal Wkly Rep. 2006 Nov 9;56(44):1157-61.
5. Thorleifsson G, Magnusson KP, Sulem P, Walters GB, Gudbjartsson DF, Stefansson H, et al. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. Science. 2007 Sep 7;317(5843):1397-400.
6. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007 Jun 8;316(5830):1491-3.
7. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet. 2007 Dec;39(12):1443-52.
8. Stefansson H, Rye DB, Hicks A, Petursson H, Ingason A, Thorgeirsson TE, et al. A genetic risk factor for periodic limb movements in sleep. N Engl J Med. 2007 Aug 16;357(7):639-47.
9. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet. 2007 May;39(5):631-7.
10. Kostulas K, Gretarsdottir S, Kostulas V, Manolescu A, Helgadottir A, Thorleifsson G, et al. PDE4D and ALOX5AP genetic variants and risk for Ischemic Cerebrovascular Disease in Sweden. J Neurol Sci. 2007 Dec 15;263(1-2):113-7.
11. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature. 2007 Jul 19;448(7151):353-7.
12. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun 7;447(7145):661-78.
13. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet. 2008 May;40(5):623-30.
14. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet. 2008 May;40(5):631-7.
15. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet. 2008 Apr;40(4):395-402.
16. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet. 2008 Feb;40(2):204-10.

17. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, et al. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet. 2008 May;40(5):575-83.
18. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, et al. Many sequence variants affecting diversity of adult human height. Nat Genet. 2008 May;40(5):609-15.
19. Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. Nat Genet. 2008 May;40(5):584-91.
20. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. Nature. 2007 Jun 7;447(7145):655-60.
21. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999 Dec;55(4):997-1004.
22. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008 May;40(5):638-45.
23. Prokopenko I, McCarthy MI, Lindgren CM. Type 2 diabetes: new genes, new understanding. Trends Genet. 2008 Oct 24.
24. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet. 2006 Jun;38(6):652-8.
25. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006 Mar;38(3):320-3.
26. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. 2008 Nov 7;322(5903):881-8.
27. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature. 1995 Dec 21-28;378(6559):789-92.
28. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001 Feb 16;291(5507):1304-51.
29. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, et al. A high-resolution recombination map of the human genome. Nat Genet. 2002 Jul;31(3):241-7.
30. A haplotype map of the human genome. Nature. 2005 Oct 27;437(7063):1299-320.
31. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. Linkage disequilibrium in the human genome. Nature. 2001 May 10;411(6834):199-204.
32. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet. 2005 Nov;37(11):1217-23.
33. Pe'er I, de Bakker PI, Maller J, Yelensky R, Altshuler D, Daly MJ. Evaluating and improving power in whole-genome association studies using fixed marker sets. Nat Genet. 2006 Jun;38(6):663-7.
34. Association AP. Diagnostic and Statistical Manual of Mental Disorders. 4th Edition, Text Revision ed. Washington, D.C.: American Psychiatric Association; 2000.

35. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. Br J Addict. 1991 Sep;86(9):1119-27.
36. Swan GE, Hops H, Wilhelmsen KC, Lessov-Schlaggar CN, Cheng LS, Hudmon KS, et al. A genome-wide screen for nicotine dependence susceptibility loci. Am J Med Genet B Neuropsychiatr Genet. 2006 Jun 5;141B(4):354-60.
37. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet. 2007 Jan 1;16(1):24-35.
38. Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. Hum Mol Genet. 2007 Jan 1;16(1):36-49.
39. Straub RE, Sullivan PF, Ma Y, Myakishev MV, Harris-Kerr C, Wormley B, et al. Susceptibility genes for nicotine dependence: a genome scan and followup in an independent sample suggest that regions on chromosomes 2, 4, 10, 16, 17 and 18 merit further study. Mol Psychiatry. 1999 Mar;4(2):129-44.
40. Fagerstrom KO, Kunze M, Schoberberger R, Breslau N, Hughes JR, Hurt RD, et al. Nicotine dependence versus smoking prevalence: comparisons among countries and categories of smokers. Tob Control. 1996 Spring;5(1):52-6.
41. Fagerstrom K, Furberg H. A comparison of the Fagerstrom Test for Nicotine Dependence and smoking prevalence across countries. Addiction. 2008 May;103(5):841-5.
42. Weiss RB, Baker TB, Cannon DS, von Niederhausern A, Dunn DM, Matsunami N, et al. A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. PLoS Genet. 2008 Jul;4(7):e1000125.
43. Martinez-Ortega JM, Jurado D, Gurpegui M. Nicotine dependence vs. daily smoking as a meaningful variable: Implications for clinical and epidemiological psychiatric studies. Prog Neuropsychopharmacol Biol Psychiatry. 2008 Sep 30.
44. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry. 1988 Dec;45(12):1069-77.
45. Madden PA, Pedersen NL, Kaprio J, Koskenvuo MJ, Martin NG. The epidemiology and genetics of smoking initiation and persistence: crosscultural comparisons of twin study results. Twin Res. 2004 Feb;7(1):82-97.
46. Saccone SF, Pergadia ML, Loukola A, Broms U, Montgomery GW, Wang JC, et al. Genetic linkage to chromosome 22q12 for a heavy-smoking quantitative trait in two independent samples. Am J Hum Genet. 2007 May;80(5):856-66.
47. Istvan J, Matarazzo JD. Tobacco, alcohol, and caffeine use: a review of their interrelationships. Psychol Bull. 1984 Mar;95(2):301-26.
48. Miller NS, Gold MS. Comorbid cigarette and alcohol addiction: epidemiology and treatment. J Addict Dis. 1998;17(1):55-66.
49. Arsrit SAA 2007. Reykjavik, Iceland: SAA2007.
50. Funk D, Marinelli PW, Le AD. Biological processes underlying co-use of alcohol and nicotine: neuronal mechanisms, cross-tolerance, and genetic factors. Alcohol Res Health. 2006;29(3):186-92.

51. Littleton J, Barron S, Prendergast M, Nixon SJ. Smoking kills (alcoholics)! shouldn't we do something about it? Alcohol Alcohol. 2007 May-Jun;42(3):167-73.
52. Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. Addiction. 2003 Jan;98(1):23-31.
53. Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. Behav Genet. 2005 Jul;35(4):397-406.
54. Kendler KS, Myers J, Prescott CA. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. Arch Gen Psychiatry. 2007 Nov;64(11):1313-20.
55. Kendler KS, Schmitt E, Aggen SH, Prescott CA. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. Arch Gen Psychiatry. 2008 Jun;65(6):674-82.
56. Kalamida D, Poulas K, Avramopoulou V, Fostieri E, Lagoumintzis G, Lazaridis K, et al. Muscle and neuronal nicotinic acetylcholine receptors. Structure, function and pathogenicity. FEBS J. 2007 Aug;274(15):3799-845.
57. Laviolette SR, van der Kooy D. Blockade of mesolimbic dopamine transmission dramatically increases sensitivity to the rewarding effects of nicotine in the ventral tegmental area. Mol Psychiatry. 2003 Jan;8(1):50-9, 9.
58. Kalivas PW. Neurotransmitter regulation of dopamine neurons in the ventral tegmental area. Brain Res Brain Res Rev. 1993 Jan-Apr;18(1):75-113.
59. Klink R, de Kerchove d'Exaerde A, Zoli M, Changeux JP. Molecular and physiological diversity of nicotinic acetylcholine receptors in the midbrain dopaminergic nuclei. J Neurosci. 2001 Mar 1;21(5):1452-63.
60. Laviolette SR, van der Kooy D. The neurobiology of nicotine addiction: bridging the gap from molecules to behaviour. Nat Rev Neurosci. 2004 Jan;5(1):55-65.
61. Kaiser S, Wonnacott S. alpha-bungarotoxin-sensitive nicotinic receptors indirectly modulate [(3)H]dopamine release in rat striatal slices via glutamate release. Mol Pharmacol. 2000 Aug;58(2):312-8.
62. Koob GF, Nestler EJ. The neurobiology of drug addiction. J Neuropsychiatry Clin Neurosci. 1997 Summer;9(3):482-97.
63. Yun IA, Wakabayashi KT, Fields HL, Nicola SM. The ventral tegmental area is required for the behavioral and nucleus accumbens neuronal firing responses to incentive cues. J Neurosci. 2004 Mar 24;24(12):2923-33.
64. Berrettini W, Yuan X, Tozzi F, Song K, Francks C, Chilcoat H, et al. alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. Mol Psychiatry. 2008 Jan 29.
65. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature. 2008 Apr 3;452(7187):633-7.
66. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet. 2008 May;40(5):616-22.
67. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008 Apr 3;452(7187):638-42.

68. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. Br Med J. 1952 Dec 13;2(4797):1271-86.
69. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. J Epidemiol Community Health. 1978 Dec;32(4):303-13.
70. Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. J Am Med Assoc. 1950 May 27;143(4):329-36.
71. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin. 2005 Mar-Apr;55(2):74-108.
72. Hirsch AT, Haskal ZJ, Hertzer NR, Bakal CW, Creager MA, Halperin JL, et al. ACC/AHA 2005 Practice Guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): a collaborative report from the American Association for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (Writing Committee to Develop Guidelines for the Management of Patients With Peripheral Arterial Disease): endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation; National Heart, Lung, and Blood Institute; Society for Vascular Nursing; TransAtlantic Inter-Society Consensus; and Vascular Disease Foundation. Circulation. 2006 Mar 21;113(11):e463-654.
73. Powell JT, Edwards RJ, Worrell PC, Franks PJ, Greenhalgh RM, Poulter NR. Risk factors associated with the development of peripheral arterial disease in smokers: a case-control study. Atherosclerosis. 1997 Feb 28;129(1):41-8.
74. Price JF, Mowbray PI, Lee AJ, Rumley A, Lowe GD, Fowkes FG. Relationship between smoking and cardiovascular risk factors in the development of peripheral arterial disease and coronary artery disease: Edinburgh Artery Study. Eur Heart J. 1999 Mar;20(5):344-53.
75. Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza RA, Xuei X, et al. Variants in Nicotinic Receptors and Risk for Nicotine Dependence. Am J Psychiatry. 2008 Jun 2.
76. Spitz MR, Amos CI, Dong Q, Lin J, Wu X. The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. J Natl Cancer Inst. 2008 Nov 5;100(21):1552-6.
77. Wang JC, Grucza R, Cruchaga C, Hinrichs AL, Bertelsen S, Budde JP, et al. Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence. Mol Psychiatry. 2008 Apr 15.
78. Grucza RA, Wang JC, Stitzel JA, Hinrichs AL, Saccone SF, Saccone NL, et al. A Risk Allele for Nicotine Dependence in CHRNA5 Is a Protective Allele for Cocaine Dependence. Biol Psychiatry. 2008 May 31.
79. Lev-Lehman E, Bercovich D, Xu W, Stockton DW, Beaudet AL. Characterization of the human beta4 nAChR gene and polymorphisms in CHRNA3 and CHRNB4. J Hum Genet. 2001;46(7):362-6.
80. Duga S, Solda G, Asselta R, Bonati MT, Dalpra L, Malcovati M, et al. Characterization of the genomic structure of the human neuronal nicotinic acetylcholine receptor CHRNA5/A3/B4 gene cluster and identification of novel intragenic polymorphisms. J Hum Genet. 2001;46(11):640-8.

81. Rempel N, Heyers S, Engels H, Sleegers E, Steinlein OK. The structures of the human neuronal nicotinic acetylcholine receptor beta2- and alpha3-subunit genes (CHRNB2 and CHRNA3). Hum Genet. 1998 Dec;103(6):645-53.
82. Wada E, Wada K, Boulter J, Deneris E, Heinemann S, Patrick J, et al. Distribution of alpha 2, alpha 3, alpha 4, and beta 2 neuronal nicotinic receptor subunit mRNAs in the central nervous system: a hybridization histochemical study in the rat. J Comp Neurol. 1989 Jun 8;284(2):314-35.
83. Xu W, Orr-Urtreger A, Nigro F, Gelber S, Sutcliffe CB, Armstrong D, et al. Multiorgan autonomic dysfunction in mice lacking the beta2 and the beta4 subunits of neuronal nicotinic acetylcholine receptors. J Neurosci. 1999 Nov 1;19(21):9298-305.
84. Drago J, McColl CD, Horne MK, Finkelstein DI, Ross SA. Neuronal nicotinic receptors: insights gained from gene knockout and knockin mutant mice. Cell Mol Life Sci. 2003 Jul;60(7):1267-80.
85. Rassadi S, Krishnaswamy A, Pie B, McConnell R, Jacob MH, Cooper E. A null mutation for the alpha3 nicotinic acetylcholine (ACh) receptor gene abolishes fast synaptic activity in sympathetic ganglia and reveals that ACh output from developing preganglionic terminals is regulated in an activity-dependent retrograde manner. J Neurosci. 2005 Sep 14;25(37):8555-66.
86. Wada E, McKinnon D, Heinemann S, Patrick J, Swanson LW. The distribution of mRNA encoded by a new member of the neuronal nicotinic acetylcholine receptor gene family (alpha 5) in the rat central nervous system. Brain Res. 1990 Aug 27;526(1):45-53.
87. Salas R, Pieri F, De Biasi M. Decreased signs of nicotine withdrawal in mice null for the beta4 nicotinic acetylcholine receptor subunit. J Neurosci. 2004 Nov 10;24(45):10035-9.
88. Orr-Urtreger A, Kedmi M, Rosner S, Karmeli F, Rachmilewitz D. Increased severity of experimental colitis in alpha 5 nicotinic acetylcholine receptor subunit-deficient mice. Neuroreport. 2005 Jul 13;16(10):1123-7.
89. Ramirez-Latorre J, Yu CR, Qu X, Perin F, Karlin A, Role L. Functional contributions of alpha5 subunit to neuronal acetylcholine receptor channels. Nature. 1996 Mar 28;380(6572):347-51.
90. Dineley-Miller K, Patrick J. Gene transcripts for the nicotinic acetylcholine receptor subunit, beta4, are distributed in multiple areas of the rat central nervous system. Brain Res Mol Brain Res. 1992 Dec;16(3-4):339-44.
91. Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ, et al. Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. Psychol Med. 1997 Nov;27(6):1381-96.
92. Merikangas KR, Stolar M, Stevens DE, Goulet J, Preisig MA, Fenton B, et al. Familial transmission of substance use disorders. Arch Gen Psychiatry. 1998 Nov;55(11):973-9.
93. Krystal JH, Carter CS, Geschwind D, Manji HK, March JS, Nestler EJ, et al. It is time to take a stand for medical research and against terrorism targeting medical scientists. Biol Psychiatry. 2008 Apr 15;63(8):725-7.
94. Thorgeirsson TE, Oskarsson H, Desnica N, Kostic JP, Stefansson JG, Kolbeinsson H, et al. Anxiety with panic disorder linked to chromosome 9q in Iceland. Am J Hum Genet. 2003 May;72(5):1221-30.

95. Gulcher JR, Kristjansson K, Gudbjartsson H, Stefansson K. Protection of privacy by third-party encryption in genetic research in Iceland. Eur J Hum Genet. 2000 Oct;8(10):739-42.
96. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet. 2006 Jun;38(6):659-62.
97. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. Science. 2004 Apr 23;304(5670):581-4.
98. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, et al. Comparison of fine-scale recombination rates in humans and chimpanzees. Science. 2005 Apr 1;308(5718):107-11.
99. Buckland PR, Hoogendoorn B, Coleman SL, Guy CA, Smith SK, O'Donovan MC. Strong bias in the location of functional promoter polymorphisms. Hum Mutat. 2005 Sep;26(3):214-23.
100. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. Genetics of gene expression and its effect on disease. Nature. 2008 Mar 27;452(7186):423-8.
101. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, et al. Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet. 2008 Aug 17.