Expression Genetics in the Human Brain: Evolution and Disease

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Ryan Matthew Smith

Neuroscience Graduate Studies Program

The Ohio State University

2010

Dissertation Committee:

Wolfgang Sadee, Advisor
Christine E. Beattie
Jack A. Boulant
John D. Oberdick
Abstract

Genetic diversity is a major factor driving phenotypic differences between individuals. The wide spectrum of heritable traits ranges from mild personality differences to extreme instances of Mendelian genetic disorders. Consequently, researchers have expended significant resources attempting to find genetic variants that contribute to phenotypic diversity, especially those relating to human disease. Past efforts have been most successful at uncovering highly penetrant mutations that occur in the protein coding regions of genes. Promoter and intronic variants can also have drastic effects on phenotype, with recent estimates suggesting that variants acting at the RNA level contribute to greater than 60% of all genetic disease. Most of these genetic variants affect RNA processing, especially transcript splicing.

Here, I present a series of studies exploring relationships between genetic variants that significantly affect RNA expression and three neurological phenotypes. First, I uncovered a transcription enhancer region contributing to nicotinic α5 receptor subunit mRNA expression and subsequently tested its relationship to nicotine addiction. Next, I revealed a significant correlation between splicing of serotonin 2A receptor mRNA transcripts and a genetic variant implicated in clinical responsiveness to atypical antipsychotics and antidepressants. Finally, in a survey of autism-related risk genes, I uncovered
evidence for genetic variations affecting mRNA expression in multiple genes, including cellular adhesion molecules, alternative splicing molecules, and integral components of glutamatergic and gamma-aminobutyric acid signaling. The broad range of neurological phenotypes affected by non-protein coding variants investigated here suggests a pervasive role for RNA processing in human disease. The accumulation of these RNA processing variants throughout human evolution is discussed in the context of human disease.
Dedicated to my wife and son, grandmother and grandfather, mother, father, brother, extended family, and great friends, whose love and support has fueled my ambition. My achievements are a reflection of your sacrifices, all of which are more greatly appreciated than can be expressed through written word.
Acknowledgments

I am grateful to acknowledge my many personal and professional mentors at Ohio State. Specifically, Wolfgang Sadee, thank you for the opportunity to study as a member of your laboratory and for challenging me to be a better scientist and professional. David Beversdorf, thank you for providing my first opportunity to understand the human mind from a scientific perspective. Members of my dissertation committee, Christine Beattie, Jack Boulant, and John Oberdick, thank you for your time and support. Finally, to my many colleagues at Ohio State, thank you for contributing to my success by providing me with knowledge, companionship, and advice.
December 5th, 1981.......................... Born, Fremont, Ohio

June 2010..................................... Gibsonburg High School

2003 to 2005................................. Undergraduate Research Associate,
                                      Department of Neurology, The Ohio
                                      State University

June 12th 2005.............................. B.A. Psychology, The Ohio State
                                      University

2006 ........................................... Post-Baccalaureate Research Assistant,
                                      Department of Neurology, The Ohio
                                      State University

2006 to Present ............................. Graduate Research Associate,
                                      Departments of Neurology, Psychology,
                                      and Pharmacology, The Ohio State
                                      University

2008 .......................................... International Meeting for Autism
                                      Research Student Travel Award

2009 .......................................... Edward J. Ray Travel Award for
                                      Scholarship and Service
Vita Con’t

Publications


Vita Con’t


Field of Study

Major Field: Neuroscience Graduate Studies Program
# Table of Contents

Abstract ........................................................................................................................................... ii

Acknowledgments ....................................................................................................................... v

Vita .................................................................................................................................................. vi

List of Tables ................................................................................................................................... xi

List of Figures ............................................................................................................................... xii

Chapter 1: Introduction to Functional Genetics ............................................................................. 1

Chapter 2: Functional Genetics of *CHRNA5*: Role in Nicotine Dependence .................. 9

2.1 Introduction and Background ................................................................................................. 9

2.2 Materials and Methods ........................................................................................................ 13

2.3 Results ................................................................................................................................... 18

2.4 Discussion ............................................................................................................................. 29

Chapter 3: Variability in Serotonin 2A Receptor Genetics and Therapeutic Response ............. 34

3.1 Introduction and Background ............................................................................................... 34

3.2 Materials and Methods ........................................................................................................ 39

3.3 Results ................................................................................................................................... 42
3.4 Discussion ........................................................................................................... 48

Chapter 4: Surveying Candidate Risk Genes in Autism Spectrum Disorders..... 55

4.1 Introduction and Background ........................................................................... 55

4.2 Materials and Methods ..................................................................................... 59

4.3 Results .............................................................................................................. 62

4.4 Discussion ......................................................................................................... 69

Chapter 5: Concluding Remarks ............................................................................ 77

References .............................................................................................................. 82

Appendix A: Abbreviations ..................................................................................... 114

Appendix B: Tables .................................................................................................. 118

Appendix C: Figures .................................................................................................. 128
List of Tables

Table 1. Correlation and LD between rs3841324 and selected SNPs.............. 12
Table 2. Influence of demographic variables on CHRNA5 mRNA expression in BA46 prefrontal cortex samples......................................................... 20
Table 3. Single-SNP and joint-effect association models testing for nicotine dependence in COGEND ........................................................................... 28
Table 4. rs905740/rs16969968 haplotype frequencies in brain samples......... 28
Table 5. Influence of demographic variables on HTR2A mRNA expression in BA46 prefrontal cortex samples................................................................. 43
Table 6. Brodmann Area 10 cortex tissue demographics.............................. 61
Table 7. CHRNA5 primer sequences for quantitative AEI and qPCR assays... 119
Table 8. Loci sequenced in CHRNA5 region.............................................. 120
Table 9. SNP associations with overall CHRNA5 expression and AEI, and LD with rs905740 ...................................................................................... 121
Table 10. Linear regression for CHRNA5 mRNA expression across genotype 122
Table 11. HTR2A primers used for SNaPshot, qPCR, and cDNA synthesis .... 123
Table 12. Primers for ASD genes investigate in BA10 cortex samples ......... 126
Table 13. Expression level (C_T) and AEI for autism risk genes................. 127

xi
List of Figures

Figure 1. Modalities of functional SNPs................................................................. 5
Figure 2. LD in the CHRNA5 locus ...................................................................... 11
Figure 3. An example of AEI in a heterozygous sample ....................................... 15
Figure 4. CHRNA5 gene map and SNP locations ................................................. 20
Figure 5. AEI measurements for CHRNA5 in all brain samples ....................... 21
Figure 6. LD plot for CHRNA5 SNPs in prefrontal cortex samples ...................... 23
Figure 7. Tissue-specific CHRNA5 expression compared across rs905470 ....... 27
Figure 8. Ets-1 binding site created by rs905740 ................................................. 30
Figure 9. HTR2A gene map denoting pharmacogenomic candidate SNPs ...... 37
Figure 10. AEI at three HTR2A marker SNPs ....................................................... 44
Figure 11. AEI comparison in samples heterozygous for multiple HTR2A marker
           SNPs ........................................................................................................... 45
Figure 12. Alternative splicing of HTR2A transcripts ......................................... 46
Figure 13. Relative amount of E2\textsuperscript{fr} spliceoform compared across rs6313
           genotype....................................................................................................... 48
Figure 14. Consensus sites for alternative splicing of HTR2A E2\textsuperscript{fr} transcript ..... 50
Figure 15. E2\textsuperscript{fr} and rs6313 interaction predicted by HTR2A exon 2 folding ..... 50
Figure 16. 5-HT2A protein and predicted E2\textsuperscript{tr} truncation ............................ 52
Figure 17. Effect of mutations in cis- and trans-acting factors on RNA splicing... 57
Figure 18. AEI in SLC1A1 measured at three marker SNP locations .................. 63
Figure 19. Within-sample comparison of AEI at multiple SLC1A1 SNP positions........................................................................................................................................................................... 64
Figure 20. Significant AEI for CNTNAP2. .................................................................. 65
Figure 21. Significant AEI for HNRNPA3. ................................................................. 66
Figure 22. Significant AEI for NOVA1........................................................................ 67
Figure 23. Significant AEI for SF4 ............................................................................ 68
Figure 24. Alignment of vertebrate 5-HT2A translation start sites (TSS), exon 2 splicing, and rs6313................................................................................................................ 129
Chapter 1: Introduction to Functional Genetics

Inheritable genetic diversity is responsible for driving evolution of all living organisms, a notion put forth by Charles Darwin in his description of natural selection in *On the Origin of Species*. At the crux of evolution is the high fidelity replicator molecule, deoxyribonucleic acid (DNA). Although it has been widely accepted since the 1950’s that DNA encodes genetic information passed from parent to progeny, it was not until 2003 that the Human Genome Project fully decoded the human genome [International Human Genome Sequencing Consortium (IHGSC), 2004]. This initial effort and many subsequent contributions revealed a 3 billion base diploid genome that is up to 99.5% identical between any two human beings (IHGSC, 2004; Lander et al., 2001; Levy et al., 2007; She et al., 2004; Venter et al., 2001). Genetic diversity between humans consists of millions of single nucleotide polymorphisms (SNPs), insertions, deletions, translocations, inversions, and variable repeats of genetic material.

Well before the Human Genome Project endowed scientists with an in-depth account of genetic heritage, it was known that genetic factors drove susceptibility to some human diseases. Notable milestones demonstrating the link between genetics and disease include the discovery of a third copy of
chromosome 21 responsible for Down syndrome in 1959 (Jacobs et al., 1959), and discovery of the risk gene for Huntington’s disease, which has the distinction of being the first gene mapped to an autosome in 1983 (Gusella et al., 1983).

Through a variety of methods, genetic contributions to human phenotypes continue to be unmasked. For highly heritable monogenic disorders, familial linkage studies often provide the resolution necessary for identifying the genomic locus harboring the risk gene. Classical linkage studies were implemented by genotyping loci throughout the genome of an affected proband and other family members with or without the disease. Segregation of pathogenic regions through recombination and inheritance in the affected, but not unaffected family members, allowed researchers to narrow down putative risk loci. Follow-up studies that more precisely map variability within the immediate risk locus pinpoint pathogenic mutations. This approach has elucidated the disease-causing variants in a number of Mendelian disorders, examples of which include Rett syndrome (Anvret et al., 1985), Duchenne muscular dystrophy (Murray et al., 1982), and amyotrophic lateral sclerosis (Siddique et al., 1991).

The utility of linkage studies is limited to cases where familial transmission of a disorder is indicated and the pathogenic mutation is highly penetrant. This approach is inadequate at explaining the majority of heritable risk in sporadic and complex disorders where many low-penetrance factors converge to manifest the disease phenotype, as hypothesized for autism (Muhle et al., 2004) or schizophrenia (Allan et al., 2008). Fueled by the genetic diversity uncovered by
the Human Genome Project, genome-wide association studies (GWAS) are a popular approach for uncovering risk loci in complex genetic diseases. GWAS is typically implemented by genotyping hundreds of thousands, or even millions of SNPs in affected and control populations and associating the genotype at all of these SNP locations with the presence or absence of the disorder. The need to use millions of SNPs in GWAS in order to get adequate representation of human genetic diversity presents an issue of statistical power, since millions of hypotheses are being tested independently, requiring large patient populations and well-defined phenotypes to uncover risk loci. Though not an exhaustive list, risk genes have been identified for autism (Anney et al., 2010; Glessner et al., 2009; Pinto et al., 2010; Wang et al., 2009a; Weiss et al., 2009), schizophrenia (O’Donovan et al., 2008), type 2 diabetes (Saxena et al., 2007; Voight et al., 2010; Yamauchi et al., 2010; Zeggini et al., 2008), coronary artery disease (Erdmann et al., 2009; Samani et al., 2007; Trégouët et al., 2009), and multiple forms of drug abuse (Bierut et al., 2007; Bierut et al., 2010; Drgon et al., 2010; Nielsen et al., 2010; Thorgeirsson et al., 2010; Uhl et al., 2008), using GWAS.

Unlike familial linkage studies where a risk locus can be definitively identified in a family or single individual, the implicated loci estimated by GWAS typically contribute modest risk for a population, providing little, if any risk estimation for an individual. For example, a recent GWAS in two large populations of autistic patients found a significant odds ratio of 1.19 for a specific SNP (rs4307059) on chromosome 5 (Wang et al., 2009). In practice, this means
that individuals in the affected population are 1.19 times more likely to carry the risk genotype at this locus versus the control population. Thus, the presence of a significant GWAS signal does not necessarily imply any functional significance for rs4307059, but simply that this SNP can act as a surrogate marker for the many SNPs in high linkage disequilibrium (LD) with rs4307059 – extending over 100 kilobases – all of which are significantly associated with the very modest increase in autism risk.

Often, significant GWAS findings require further efforts to locate specific variants within the immediate genomic locus that have a biological function, begging the question of how functional variants are defined. The fidelity of protein translation based on DNA sequence allows pathogenic amino acid-encoding mutations (coding SNPs; cSNPs) to be easily predicted from the DNA sequence of affected individuals. Less apparent, and much more difficult to predict from genomic sequence, are functional variants that affect RNA transcription (regulatory SNPs; rSNPs) or heterogeneous nuclear RNA (hnRNA) processing, including RNA folding, alternative splicing, polyadenlyation, microRNA binding, etc. (structural RNA SNPs; srSNPs). Figure 1 highlights important characteristics of cSNPs, rSNPs, and srSNPs.
Modalities of functional SNPs. cSNPs are non-synonymous variants that change the protein-coding sequence of a gene, exerting their effect on all proteins made from the affected gene. rSNPs typically change transcription factor binding sites in promoter or enhancer regions to affect overall transcription of a gene. srSNPs can be present in introns, exons, or 5’ and 3’ untranslated regions, modulating the processing of heterogeneous nuclear RNA. Affected processes can include splicing, turnover, folding, and transcriptional or translational efficiency.

In contrast to single-gene disorders, such as cystic fibrosis, where a majority of patients are homozygous for a common pathogenic cSNP (Cutting et al., 1989), the search for functional cSNPs in complex genetic disorders has yielded few pathogenic mutations that are shared amongst all affected individuals. This missing heritability prompted researchers to reevaluate the nature of pathogenic mutations in complex disorders, where they stressed the importance of searching outside of coding regions for disease-causing variants.
(Manolio et al., 2009). When intronic and exonic regions of risk genes are studied in a systematic manner, approximately 50% of disease-causing variants reside outside of coding regions and affect RNA processing (Ars et al., 2000; Cartegni et al., 2002; Mayer et al., 1999; Mayer et al., 2000; Messiaen et al., 2000; Pagenstecher et al., 2006; Teraoka et al., 1999). Global estimates based on mathematical models suggest up to 60% of all disease-causing mutations disrupt RNA processing (Lopez-Bigas et al., 2005; Wang and Cooper, 2007) and likely reside in intronic regions. Thus, identifying functional rSNPs and srSNPs within implicated risk genes has the potential to fill some of the missing heritability associated with complex genetic disorders. A meta-analysis of GWAS studies note that 88% of trait-associated SNPs are located outside of protein-coding regions of genes (Hindorff et al., 2009). It is further hypothesized that these trait-associated SNPs, especially those residing in intergenic regions, regulate RNA transcription (Hindorff et al., 2009). Genome-wide efforts to identify functional variants responsible for modulating gene transcription have taken shape in the form of expression quantitative trait loci analyses (eQTL). By combining microarray platforms that measure expression and genotype SNPs on a genome-wide basis, the relationship between RNA expression and genotype can be elucidated, in a manner similar to GWAS. The use of this technique has identified numerous loci responsible for modulating gene expression in human cortex (Myers et al., 2007). While useful, the eQTL approach is subject to the same limitations as GWAS, most notably the inability to account for trans-acting
factors that increase in proportion to the heterogeneity of the population under study. This includes environmental factors that can differ between populations and are difficult to recognize and take into account. Another drawback to eQTL analysis is the inability to distinguish between genetic variants in high LD confounding efforts to locate the functional variant truly responsible for the observed differences in expression.

The drawbacks inherent to eQTL analysis can be alleviated by measuring allelic expression of mRNA transcripts. By measuring the amount of expression between two alleles in the same individual, the influence of trans-acting factors is eliminated, since the trans-acting factor would have equal opportunity to act on each allele from the same individual. Thus, an allelic expression phenotype can be constructed for each RNA transcript isoform, even in heterogeneous populations. The allelic expression phenotype can then be used to scan the genomic region of interest for SNPs that are responsible for the allelic expression differences, even in the presence of high LD. Previous studies using this approach have been successful at uncovering cis-regulatory variants (Johnson et al., 2008a; Wang et al., 2006), or cis-eQTLs (Ge et al., 2009). Advances in massive-parallel sequencing technologies now allow this approach to be employed on a genome-wide basis (Heap et al., 2010).

Following is a series of studies aimed at uncovering functional SNPs within risk genes implicated in a range of phenotypes. The first study demonstrates the presence of an enhancer region harboring six SNPs that
resides upstream of nicotinic α5 receptor subunit gene (CHRNA5), the minor alleles of which correspond to a significant increase in CHRNA5 mRNA expression and risk for nicotine dependence. The second study correlates alternative splicing of serotonin 2A receptor gene (HTR2A) mRNA with a SNP that predicts clinical outcomes for individuals treated with antidepressants and atypical antipsychotic drugs. Finally, a survey of autism-related risk genes provides evidence for the presence of common genetic variants that significantly affect mRNA expression within cellular adhesion molecules, components of glutamatergic and gamma-aminobutyric acid (GABA) signaling, and alternative splicing proteins. Importantly, all of these studies lend support to the notion that rSNPs and srSNPs, occurring at a high frequency in the general population, contribute to human phenotypic diversity and disease.
Chapter 2: Functional Genetics of CHRNA5: Role in Nicotine Dependence

2.1 Introduction and Background

Risk for nicotine addiction is partially driven by genetic factors. Assessments of heritability using twin and adoption studies suggest at least 50% of liability is genetic (Li, 2006). Increased health care costs for young cigarette smokers (Barendregt et al., 1997) and substantial decreases in life expectancy (Doll et al., 2004) make smoking prevention a financial and moral imperative. Understanding genetic factors contributing to smoking addiction can inform researchers about biological mechanisms underlying dependence, allowing development of more effective prevention and cessation programs.

Genetic association studies link the nicotinic receptor subunit α5 gene (CHRNA5) to smoking (Berrettini et al., 2008; Bierut et al., 2008; Saccone et al., 2007; Saccone et al., 2009a and 2009b; Sherva et al., 2008 and 2010; Stevens et al., 2008; Tobacco and Genetics Consortium, 2010; Wang et al., 2009b; Weiss et al., 2008), lung cancer (Amos et al., 2008; Hung et al., 2008; Liu et al., 2010; Wang et al., 2009b), chronic obstructive pulmonary disease (COPD) (Pillai et al., 2009), cocaine addiction (Grucza et al., 2008; Sherva et al., 2010), and alcohol dependence (Schlaepfer et al., 2008; Sherva et al., 2008; Wang et al., 2009c). In individuals of European descent, CHRNA5 harbors only one frequent
non-synonymous cSNP, rs16969968 in exon 5 (minor allele frequency ~40%), resulting in an aspartic acid to asparagine residue change (D398N). This variant is explicitly associated with risk for nicotine dependence, lung cancer, and COPD (Amos et al., 2008; Bierut et al., 2008; Hung et al., 2008; Liu et al., 2010; Pillai et al., 2009; Saccone et al., 2007; Saccone et al., 2009a and 2009b; Sherva et al., 2008 and 2010; Stevens et al., 2008; Tobacco and Genetics Consortium, 2010; Wang et al., 2009b and 2009c; Weiss et al., 2008). In vitro, rs16969968 reduces the maximal response of the receptor to the agonist epibatidine (Bierut et al., 2008), while response to nicotine and acetylcholine remains to be determined.

Expression of α5 mRNA appears to be modulated by genetic factors. Similar to genetic association studies where a genotype is associated with a clinical phenotype, gene expression can be substituted as a quantitative phenotype, and the genome probed for regions that predict expression. Regions correlating with expression are termed expression quantitative trait loci (eQTL). To uncover eQTLs within CHRNA5, two previous studies correlated mRNA expression in post-mortem prefrontal cortex tissue to CHRNA5 polymorphisms, yielding significant associations between α5 expression and a 22-base pair deletion (rs3841324) located less than 100 bases upstream of the CHRNA5 transcription start (Wang et al., 2009b and 2009c), proposed to function as a promoter variant. Homozygous carriers of the variant short allele expressed 2- to 2.9-fold more CHRNA5 mRNA versus the wild-type long allele in human prefrontal cortex (Wang et al., 2009b and 2009c).
One major drawback to the eQTL method is its inability to distinguish between loci that are in high LD, in order to correctly identify the region responsible for driving differences in expression. Due to lengthy regions of strong LD in the α5 region (Figure 2), where multiple other SNP genotypes are correlated with rs3841324 genotype (Table 1), the association between rs3841324 and mRNA expression levels in brain is less than definitive. In addition, total mRNA levels used to determine eQTLs are affected by multiple trans-acting factors as well as variants in cis-regulatory regions.

Figure 2. LD in the CHRNA5 locus. Solid red corresponds to complete LD ($D'=1$).
### Table 1. Correlation and LD between rs3841324 and selected SNPs across the \textit{CHRNA5} gene locus in samples of European ancestry.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Correlation ((R))</th>
<th>LD ((D'))</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1979907</td>
<td>0.871193</td>
<td>0.911679</td>
</tr>
<tr>
<td>rs880395</td>
<td>0.895478</td>
<td>0.91656</td>
</tr>
<tr>
<td>rs905740</td>
<td>0.894711</td>
<td>0.917173</td>
</tr>
<tr>
<td>rs7164030</td>
<td>0.894711</td>
<td>0.917173</td>
</tr>
<tr>
<td>rs2036527</td>
<td>0.475209</td>
<td>0.891988</td>
</tr>
<tr>
<td>rs667282</td>
<td>0.332054</td>
<td>0.648598</td>
</tr>
<tr>
<td>rs17486195</td>
<td>0.498187</td>
<td>0.999688</td>
</tr>
<tr>
<td>rs680244</td>
<td>0.796958</td>
<td>0.82443</td>
</tr>
<tr>
<td>rs621849</td>
<td>0.83532</td>
<td>0.838858</td>
</tr>
<tr>
<td>rs692780</td>
<td>0.744119</td>
<td>0.812606</td>
</tr>
<tr>
<td>rs17408276</td>
<td>0.683924</td>
<td>0.780498</td>
</tr>
<tr>
<td>rs16969968</td>
<td>0.517779</td>
<td>0.999645</td>
</tr>
<tr>
<td>rs514743</td>
<td>0.768621</td>
<td>0.788763</td>
</tr>
<tr>
<td>rs615470</td>
<td>0.698081</td>
<td>0.726942</td>
</tr>
<tr>
<td>rs578776</td>
<td>0.211199</td>
<td>0.410349</td>
</tr>
</tbody>
</table>

To mitigate trans-acting factors and distinguish between functional and marker SNPs embedded in a haplotype block with high but incomplete LD, we measured allelic mRNA expression of \textit{CHRNA5} in human brain autopsy tissues. Different mRNA expression between two alleles in the same tissue, termed allelic expression imbalance (AEI), is the most immediate and quantitative phenotype one can exploit to scan a gene locus for the presence of cis-regulatory variants (Johnson et al., 2008a; Wang et al., 2006), or cis-eQTLs (Ge et al., 2009), with greater confidence than is possible with use of total mRNA levels. This approach revealed a distal enhancer region more than 13 kb upstream of \textit{CHRNA5}, harboring six SNPs in complete LD with each other in European- and African-Americans, but in lower LD with other SNPs previously suspected of affecting expression. The minor alleles of these proposed enhancer SNPs correspond to
a 4.1-fold increase in overall \textit{CHRNA5} mRNA expression. In light of these findings, we discuss genetically variable \textit{CHRNA5} expression in disease associations and assess risk for nicotine addiction through the use of previous large scale association studies (Saccone et al., 2007; Saccone et al., 2009a and 2009b).

\section*{2.2 Materials and Methods}

\textit{Nucleic acid isolation.} Genomic DNA (gDNA) was isolated from 91 prefrontal cortex tissues (provided by Dr. Deborah Mash) originating from Brodmann Area 46, using a “salting out” method (Miller et al., 1988) supplemented with additional sodium dodecyl sulfate for lipid-rich brain tissue. Total RNA was isolated with RNeasy Mini Kit spin columns (Qiagen, Germantown, MD) from brain homogenates dissolved in TRIzol (Life Technologies, Carlsbad, CA). RNA preparations were treated with DNase to remove any remaining gDNA fragments. Two separate cDNA preparations were made using 0.5\textmu{g} total RNA per sample preparation. Oligo-dT was used to prime mRNA in the reverse transcription reaction. Similarly, gDNA and RNA from 87 patients was isolated from lymphocytes originating from the CEPH Coriell cell collection (Coriell Institute, Camden, NJ).

\textit{Brain and lymphocyte sample genotyping.} SNPs rs16969968 and rs615740, used as markers for measuring AEI, were genotyped by restriction
fragment length polymorphism (RFLP) methods. Restriction enzymes Taqα1 (rs16969968) or CviQI (rs615470) (New England Biolabs - NEB, Ipswich, MA) were used to cut wild-type alleles for gDNA amplified using primers tagged with a fluorophore (6-FAM or HEX), the fragments of which were separated via capillary electrophoresis and detected by an ABI 3730 DNA Analyzer (Life Technologies). RFLP was also used to genotype rs17486195 (Nsil), rs680244 (BsmAI), rs17408276 (CviQI), and rs578776 (NlaIII). The deletion polymorphism rs3841324 was genotyped using PCR with fluorescent primers and separated via capillary electrophoresis using an ABI 3730. SNaPshot Multiplex System (Life Technologies) was used to genotype additional SNPs: rs1979907, rs880395, rs905740, rs7164030, rs2036527, rs667282, rs621849, rs692780, and rs514743. Briefly, gDNA extracted from the tissue samples are PCR-amplified using primers targeting the SNP of interest. Unincorporated deoxyribonucleotides (dNTPs) and excess primers from the PCR reaction are subsequently degraded during a 37°, 3-hour incubation in exonuclease I and bacterial Antarctic phosphatase (New England Biolabs Inc.). Following, the treated PCR product is subjected to the primer extension SNaPshot reaction, which incorporates a fluorescent dideoxynucleotide (ddNTP) into the SNP position. Unincorporated fluorescent ddNTPs are degraded by calf intestinal phosphatase (New England Biosystems Inc.). The alleles are resolved using an ABI 3730 DNA analyzer.

**Allelic mRNA Expression Imbalance in human brain tissue.** Allele-specific mRNA expression was quantified in samples heterozygous for either or both
marker SNPs (rs16969968 and rs615470) using a primer extension assay (SNaPshot), whereby a fluorescent dideoxynucleotide (ddNTP) is incorporated into the marker SNP position. Each ddNTP nucleotide is tagged with a different colored fluorophore, allowing each tagged allele to be distinguished when samples are heterozygous at the marker SNP location (Figure 3).

![Figure 3](image)

Figure 3. An example of AEI in a heterozygous sample, measured using SNaPshot for both genomic DNA and cDNA. Note: aG = G allele; aA = A allele.

Primer sequences for quantitative measures (AEI and qPCR) are listed in Appendix B: Table 7. Because of low cortical CHRNA5 expression, 12.5ng cDNA/25ng gDNA were first pre-amplified over 10 PCR cycles using tailed primers targeting the exon containing the marker SNP, enabling equal application to both cDNA and gDNA. Subsequently, 1µl of pre-amplified cDNA/gDNA was PCR amplified 30 cycles using primers targeting the tails attached to the pre-amplification primers. Following the SNaPshot primer extension step and separation on an ABI 3730, area under the curve was determined using GeneMapper 4.0 software (Life Technologies) to calculate relative allelic
expression ratios. For each SNP, samples were measured in triplicate for each of two independent cDNA syntheses, while gDNA was measured in duplicate. Allelic ratios for cDNA were normalized against the overall average ratio obtained for gDNA for each marker SNP (ancestral/variant allele). The percent change from gDNA ratio was log_{10} transformed and used for data analysis.

**Sequencing.** Genomic DNA (25ng) from two tissues was amplified with Phusion High-Fidelity DNA Polymerase (NEB) using primers targeting 13 loci within and around the CHRNA5 gene (Appendix B: Table 8). PCR product was gel-purified using a Qiaquick Gel Extraction Kit (Qiagen) or GenElute Gel Extraction Kit (Sigma-Aldrich, St. Louis, MO) and then sequenced by the Ohio State University Plant-Microbe Genomics Facility using an ABI 3730 DNA Analyzer (Life Technologies). Sequence data were aligned using the Basic Local Alignment Search Tool – bl2seq ([http://blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)) and compared against Reference Sequence (NCBI Build 36.1 on the UCSC Genome Browser (Kent et al., 2002): [http://genome.ucsc.edu](http://genome.ucsc.edu)).

**Quantitative real-time PCR (qPCR).** CHRNA5 mRNA expression was measured in triplicate via qPCR on an ABI 7000 Sequence Detection System (Life Technologies). 12.5ng of cDNA was amplified in a 15µl reaction using Power SYBR Green PCR Master Mix (Life Technologies). Primers for CHRNA5 measurements targeted the same region as used for SNaPshot analysis of rs16969968. The cycle threshold (C_T) was used to calculate relative quantity of CHRNA5 mRNA, normalized within each sample to β-actin (ACTB) expression.
(measured in duplicate), which was used for data analysis. Nicotinic receptor subunit β4 (CHRNB4) mRNA was measured by a single qPCR reaction performed for each sample, also normalized to ACTB.

Clinical association design and sample genotyping. Nicotine dependent and non-dependent smoking subjects were recruited by the multi-site Collaborative Genetic Study of Nicotine Dependence (COGEND) project. For the current study, only rs880395 and rs16969968 were considered for association analysis, and a total of 2,050 subjects of European descent had complete genotyping data at both SNPs (1055 nicotine dependent cases and 995 non-dependent smoking controls). This cohort is described in previous genetic association studies on nicotine dependence, which also includes more detailed genotyping methods (Saccone et al., 2007; Saccone et al., 2009a and 2009b). To meet criteria for nicotine dependence, individuals must have scored greater than 4 (out of 10) on the Fagerström test for nicotine dependence (FTND), while control subjects must have smoked greater than 100 cigarettes in their lifetime, but have an FTND score of 0.

Data Analysis. Samples with allelic expression ratios >2 were considered as potentially harboring functional cis-acting polymorphisms, with a 2.5-fold allelic difference chosen as a stringent cutoff. HelixTree software (Golden Helix, Bozeman, MT) was used for testing genotype associations with allelic expression (performed for tissues heterozygous for the two marker SNPs), and performing linear regression analysis of genotype on overall CHRNA5 mRNA expression
Univariate analysis of variance comparing CHRNA5 mRNA expression across demographic categories, with age and post-mortem interval as covariates, and subsequent correlations and one-way ANOVAs comparing CHRNA5 and CHRNA4 expression across rs905740 genotype for brain and lymphocyte samples was performed with SPSS v17.0 (SPSS, Inc., Chicago, IL). SAS v9.1 (SAS Institute, Cary, NC) was used for logistic regression analyses and to calculate odds ratios for nicotine dependence risk according to genotype.

2.3 Results

Allelic expression imbalance analysis. Table 2 displays sample demographics for post-mortem prefrontal cortex (Brodmann area 46) samples used in this study, obtained from a population of cocaine abusers and controls. Cocaine use, race, sex, smoking history, age, or postmortem interval had no effect on overall CHRNA5 mRNA expression (ANOVA; Table 2). Forty-three of 91 post-mortem brain samples were heterozygous for one or both marker SNPs, with locations shown in Figure 4, yielding reproducible allelic mRNA measurements for analysis. Within each assay, standard deviations for gDNA ratios, normalized to 1 for subsequent comparisons to allelic mRNA ratios, varied by 8.9-9.8% of the mean for both marker SNPs. One sample displaying an allelic gDNA ratio greater than two standard deviations below the mean across multiple replicates was excluded from further analysis due to potential copy number...
variation. Because of the low abundance of \textit{CHRNA5} transcripts, the within sample standard deviation of the allelic mRNA ratios were higher than those for the gDNA ratios (S.D. = 26% within-subject variation for rs16969968, and 47% for rs615470). Despite the relatively high error observed for mRNA measurements, allelic mRNA ratios were consistent between two separate cDNA syntheses (correlation coefficient, \( r = 0.62 \)) and between the two marker SNPs in subjects heterozygous for both (Fig. 2; \( r = 0.73 \)). Given the large allelic mRNA expression imbalance (AEI) in multiple subjects, with ratios exceeding fourfold difference between the alleles in many samples, the S.D. of the allelic mRNA ratio assay was acceptable for detecting large AEI. We selected a conservative cutoff of allelic mRNA ratios >2.5 as indicating the presence of significant AEI, detectable with high confidence. Therefore, we cannot exclude the possibility that smaller allelic differences (ratios <2.5) are also present that escape reliable detection using this approach.

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Group (n)</th>
<th>\textit{CHRNA5} Expression (C\textsubscript{T} Avg. ± St. Dev.)</th>
<th>ANOVA (( p ))</th>
<th>No. Samples in AEI Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male (62)</td>
<td>13.4 ± 1.5</td>
<td>0.597</td>
<td>29 (18 AEI positive)</td>
</tr>
<tr>
<td></td>
<td>Female (29)</td>
<td>13.8 ± 1.4</td>
<td></td>
<td>13 (9)</td>
</tr>
<tr>
<td></td>
<td>Caucasian (59)</td>
<td>13.6 ± 1.3</td>
<td>0.904</td>
<td>28 (17)</td>
</tr>
<tr>
<td></td>
<td>African American (14)</td>
<td>13.2 ± 1.6</td>
<td></td>
<td>4 (2)</td>
</tr>
<tr>
<td></td>
<td>Mixed/Other (18)</td>
<td>13.7 ± 2.0</td>
<td></td>
<td>10 (8)</td>
</tr>
<tr>
<td>Race</td>
<td>Cocaine (39)</td>
<td>13.5 ± 1.6</td>
<td>0.809</td>
<td>21 (15)</td>
</tr>
<tr>
<td></td>
<td>Control (52)</td>
<td>13.6 ± 1.4</td>
<td></td>
<td>21 (12)</td>
</tr>
<tr>
<td>Drug Group</td>
<td>Smoker (35)</td>
<td>13.7 ± 1.6</td>
<td>0.187</td>
<td>12 (6)</td>
</tr>
<tr>
<td></td>
<td>Non-smoker (56)</td>
<td>13.4 ± 1.4</td>
<td></td>
<td>30 (21)</td>
</tr>
<tr>
<td>Smoking History</td>
<td></td>
<td></td>
<td>0.224</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>0.828</td>
<td></td>
</tr>
<tr>
<td>Postmortem Interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Influence of demographic variables on *CHRNA5* mRNA expression in BA46 prefrontal cortex samples

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1979907*</td>
<td>1323405</td>
</tr>
<tr>
<td>rs988395*</td>
<td>1323405</td>
</tr>
<tr>
<td>rs905740*</td>
<td>1323405</td>
</tr>
<tr>
<td>rs1979906*</td>
<td>1323405</td>
</tr>
<tr>
<td>rs1979905*</td>
<td>1323405</td>
</tr>
<tr>
<td>rs2036527</td>
<td>1323405</td>
</tr>
<tr>
<td>rs3841324</td>
<td>1323405</td>
</tr>
<tr>
<td>rs667282</td>
<td>1323405</td>
</tr>
<tr>
<td>rs692780</td>
<td>1323405</td>
</tr>
<tr>
<td>rs680244</td>
<td>1323405</td>
</tr>
<tr>
<td>rs17408276</td>
<td>1323405</td>
</tr>
<tr>
<td>rs514743</td>
<td>1323405</td>
</tr>
<tr>
<td>rs16969968*</td>
<td>1323405</td>
</tr>
<tr>
<td>rs615470*</td>
<td>1323405</td>
</tr>
</tbody>
</table>

Figure 4. *CHRNA5* gene map and SNP locations. *Note:* # = putative enhancer SNPs; * = marker SNPs used to measure AEI.

Considering AEI ratios with >2.5-fold difference, 27 of the 42 samples heterozygous for at least one marker SNP displayed significant AEI (Figure 5). Of the 22 samples heterozygous for rs16969968, 11 displayed significant AEI, while 25 of 31 samples heterozygous for rs615470 were AEI-positive. All significant AEI ratios showed greater expression for the ancestral allele versus the variant allele of the respective marker SNPs, suggesting the presence of a functional polymorphism that is in high LD with the two marker SNPs, as equilibrated SNPs affecting allelic expression would cause AEI ratios above and below unity. The marker SNPs themselves are unlikely to account for the AEI ratios, since several samples were heterozygous at these SNP locations but lacked AEI. In a separate pilot study of prefrontal cortex autopsy tissues obtained from the Buckeye Brain Bank at The Ohio State University, a similar frequency of large AEI ratios was observed (data not shown), supporting the accuracy of these results.
Figure 5. AEI measurements for CHRNA5 in all brain samples with at least one heterozygous marker SNP. *Note:* dotted lines = 2.5-fold difference in allelic expression.

**CHRNA5 SNP scanning and sequencing.** Using significant AEI as a binary phenotype (*i.e.* AEI-positive or AEI-negative), we performed SNP scanning across the gene locus to identify cis-eQTL regions associated with AEI. The presence of AEI suggests that there is a *cis*-acting genetic factor present on one allele, but not on the sister allele. Thus, it is expected that any SNP driving differences in expression between the two alleles are going to be heterozygous in samples where AEI is significant. Similarly, samples not displaying significant differences in expression across the two alleles are expected to be homozygous at the functional SNP position. Guided by strong blocks of LD in CHRNA5, samples were first genotyped for 10 variants along the length of the gene (Figure 4) in addition to the two marker SNPs. Several SNPs were significantly associated with AEI ratios and overall expression, shown in Appendix B: Table 9,
but no single variant fully accounted for the observed AEI, as these variants were either homozygous when AEI was present or heterozygous when it was not, or both, in at least one subject.

To identify functional polymorphisms, two samples sharing a similar haplotype (concordant genotypes at 11 of 12 polymorphic sites) but differing in AEI phenotype (one AEI positive, one negative) were sequenced at 13 loci within and surrounding CHRNA5 (Appendix B: Table 8). Polymorphisms heterozygous in the AEI-positive sample but homozygous in the AEI-negative sample were considered candidates for further study. Of 39 polymorphic sites with appreciable frequency in CEU or YRI HapMap populations (Appendix B: Table 8), only three SNPs matched these criteria. These three SNPs, rs880395, rs905740, and rs7164030, reside in a span of 300 nucleotides approximately 13.5kb upstream of the CHRNA5 transcription start site (Figure 4), and their minor allele frequency (MAF) varies by race (CEU=0.37, YRI=0.18). An additional 12 samples were sequenced at the upstream locus (Appendix B: Table 8; chromosomal position 78844131 – 78845361), confirming heterozygosity in AEI-positive samples and homozygosity in AEI-negative samples for all three SNPs. Finally, the three SNPs were genotyped in all brain tissues, showing them to be in perfect LD ($D'=1, r^2=1$) with each other and accounting fully for all AEI data. This finding applies to the population as a whole and to each race cohort. We can also exclude the possibility that any of the other genotyped SNPs are functional in other racial groups, as at least one tissue was homozygous in AEI-
positive or heterozygous in AEI-negative samples in each group. As predicted, the putative functional SNPs were in high LD with the marker SNPs used to measure AEI in our samples ($D' = 0.89$ for rs16969968, 0.70 for rs615470; Figure 6).

![LD plot for CHRNA5 SNPs in prefrontal cortex samples used to measure AEI, regression, and total mRNA expression.](image)

The SNP Annotation and Proxy Search website ([http://www.broadinstitute.org/mpg/snap/](http://www.broadinstitute.org/mpg/snap/)) (Johnson et al., 2008b) was used to find additional SNPs predicted to be in complete LD with rs880395, rs905740, or rs7164030 in both CEU and YRI HapMap populations, yielding three additional SNPs (rs1979905, rs1979906, and rs1979907) located an additional 2 kb
upstream of the newly identified SNPs and 15kb upstream of *CHRNA5*. Genotyping rs1979907 as a surrogate for all three additional SNPs confirmed complete LD between all of the upstream SNPs in our samples (Appendix B: Table 9). Furthermore, all samples with >2.5-fold change in AEI were heterozygous at these four measured SNP locations, while all samples with <2.5-fold AEI were homozygous, designating one or more of these six SNPs as the likely functional variant(s); however, we cannot exclude the possibility of additional SNPs in complete LD in more distal regions. Genotype association with AEI ratios for these SNPs reached strong significance (*p*=5.37 x 10^{-12}, Appendix B: Table 9), while other SNPs previously suspected of being responsible for altered expression, in particular rs3841324, failed to fully account for high AEI ratios and displayed lower association with AEI (*p*=0.00448, Appendix B: Table 9). Of the 42 subjects analyzed, 3 subjects showed significant AEI but were homozygous for rs3841324, while an additional 3 were heterozygous for this promoter variant and lacked AEI.

*CHRNA5 mRNA expression in brain and lymphocytes.* Linear regression was performed for the 16 polymorphisms genotyped in all 91 samples to uncover possible eQTLs, which are also subject to trans-acting factors, in contrast to the cis-eQTLs detected by AEI analysis. The results support a strong association between the putative functional variants, rs1979907, rs880395, rs905740, and rs7164030, and total brain *CHRNA5* mRNA expression (*p*=2.42 x 10^{-5}, Appendix B: Table 9). These upstream SNPs scored better than the 22-bp deletion
rs3841324 previously associated with mRNA expression ($p=1.81 \times 10^{-3}$), consistent with the AEI results. An intronic SNP, rs621849, was also highly correlated with total mRNA levels ($p=3.45 \times 10^{-6}$) but to a lesser extent with AEI ($p=7.16 \times 10^{-4}$ versus $p=5.37 \times 10^{-12}$ for the enhancer variants). The high eQTL value for rs621849 could have been a fortuitous result of trans-acting factors in the tissues analyzed, or it is possible that rs621849 is in partial LD with other regulatory variants that have lesser effects on mRNA expression not deemed significant by our AEI analysis. Also significantly associated with expression, rs3841324 and rs621849 are in substantial LD with the six upstream SNPs in our samples (Appendix B: Table 9, Figure 6), thereby at least partially accounting for their significant associations with expression. While we cannot discount the possibility that they have smaller effects (<2.5-fold) on mRNA expression, examining regression and LD within each racial cohort argues against a functional role for rs3841324. In African-Americans, LD with the upstream ‘enhancer’ SNPs is greatly decreased compared to European-American ($r^2=0.12$; Appendix B: Table 10) and the relationship between mRNA expression and rs3841324 genotype decreases accordingly ($p=0.672$; Appendix B: Table 10). Similarly, race was not a significant factor when included as a covariate in the regression analysis (Appendix B: Table 10), suggesting the upstream SNPs dominate the genetic effect on mRNA expression.

Comparing expression in brain samples across rs905740 genotype (Figure 7a) revealed a significant 4.1-fold increase of CHRNA5 expression for
homozygous carriers of the minor SNP “T” allele compared to homozygous carriers of the major wild-type “C” allele (overall ANOVA $F=9.97$, $p=0.0001$), with heterozygotes in between. As rs3841324 is in substantial LD with the upstream enhancer SNPs identified here, we find a lesser but still significant 2.9-fold increase in expression for carriers of the deletion, identical to previously reported changes in expression correlated to this variant (Wang et al., 2009b). Nevertheless, these results indicate that rs3841324 can only be considered a surrogate marker for the upstream enhancer SNPs, and only in European-Americans.

To test for tissue specific regulation, $CHRNA5$ mRNA expression was measured in human peripheral lymphocytes, showing that rs905740 genotype had no detectable effect ($F=1.83$, $p=0.17$; Figure 7b). We further considered whether the distal enhancer SNPs affect expression of the adjacent nicotine receptor subunit $CHRNB4$, because tissue-specific expression of $CHRNA5$ and $CHRNB4$ is correlated according to microarray data from the Genomics Institute of the Novartis Research Foundation (http://symatlas.gnf.org) (Su et al., 2002), ($r=0.76$ for $CHRNA5$ reporter 206533_at and $CHRNB4$ reporter 207516_at). Using RT-PCR, we confirmed a modest correlation between $CHRNA5$ and $CHRNB4$ mRNA expression in our brain samples ($r=0.58$); however, $CHRNB4$ mRNA expression did not significantly differ across rs905740 genotype ($F=0.55$, $p=0.58$).
Figure 7. Tissue-specific *CHRNA5* expression compared across rs905470 genotype measured in a) prefrontal cortex and b) peripheral blood lymphocytes.

**Enhancer SNP association with nicotine dependence.** Genotyping results generated by previous studies were used to test associations of the upstream enhancer region to nicotine dependence in European-Americans recruited by the Collaborative Genetic Study of Nicotine Dependence (COGEND) (Saccone et al., 2007; Saccone et al., 2009a and 2009b). Among the upstream enhancer SNPs, only rs880395 was included in the genotyping arrays, and when analyzed alone, failed to show any significant associations with nicotine dependence, whereas the nonsynonymous rs16969968 variant conferred significant risk (Bierut et al., 2008; Saccone et al., 2007; Saccone et al., 2009b) (Table 3). A joint SNP analysis of rs880395 and rs16969968 finds a modest, but significant contribution of rs880395 to nicotine dependence in European-Americans [odds ratio (O.R.) =1.22, confidence interval (C.I.): 1.04-1.44, \( p =0.01 \), Table 3], while the
nonsynonymous SNP rs16969968 remains robustly associated [O.R.=1.57, C.I.: 1.33-1.86, \( p=1.2 \times 10^{-7} \), Table 3]. The distribution of rs16969968 and rs880395 is non-random. The risk allele for rs16969968 (A genotype) occurs almost exclusively on the low expression main (non-risk) G allele of rs880395 (Table 4).

<table>
<thead>
<tr>
<th>Risk Allele</th>
<th>N</th>
<th>OR (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single SNP Analysis(^1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs880395</td>
<td>A</td>
<td>2050</td>
<td>0.93 (0.82-1.06)</td>
</tr>
<tr>
<td>rs16969968</td>
<td>A</td>
<td>2050</td>
<td>1.39 (1.21-1.58)</td>
</tr>
<tr>
<td>Joint SNP Analysis(^2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs880395</td>
<td>A</td>
<td>2050</td>
<td>1.22 (1.04-1.44)</td>
</tr>
<tr>
<td>rs16969968</td>
<td>A</td>
<td>2050</td>
<td>1.57 (1.33-1.86)</td>
</tr>
</tbody>
</table>

\(^1\) Logistic regression model per SNP where CASE=SNP+SEX+AGE.

\(^2\) Logistic regression model for both SNPs where CASE=SNP\(_1\)+SNP\(_2\)+SEX+AGE.

Table 3. Single-SNP and joint-effect association models testing for nicotine dependence in COGEND

<table>
<thead>
<tr>
<th>Haplotype Makeup</th>
<th></th>
<th></th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs905740 Genotype</td>
<td>rs16969968 Genotype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td></td>
<td>38%</td>
</tr>
<tr>
<td>G</td>
<td>A</td>
<td></td>
<td>29%</td>
</tr>
<tr>
<td>A</td>
<td>G</td>
<td></td>
<td>32%</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td></td>
<td>1%</td>
</tr>
</tbody>
</table>

Table 4. rs905740/rs16969968 haplotype frequencies in brain samples
2.4 Discussion

We have identified six CHRNA5 SNPs (rs1979905, rs1979906, rs1979907, rs880395, rs905740, and rs7164030), in perfect LD, located in a region more than 13 kb upstream of the transcription site that are robustly associated with CHRNA5 mRNA expression. Quantitative analysis in prefrontal cortex tissues revealed a fourfold increase in CHRNA5 mRNA expression for homozygous carriers of the minor allele enhancer SNPs, supporting the hypothesis that the six upstream SNPs are located in a highly penetrant enhancer region. Given the level of precision attainable for measuring allelic mRNA ratios of transcripts with low expression, we cannot address here whether additional regulatory polymorphisms exist that have lesser effects on expression.

Evolutionary changes in gene expression across species commonly result from polymorphisms in cis-regulatory elements that create or abolish cis-acting transcription enhancer/suppressor factor binding sites (Wittkopp et al., 2004). The accumulation of the proposed enhancer SNPs to very high allele frequencies (37% in populations of European descent), resulting in gain of function compared to the ancestral alleles, together with their position within a large haplotype block, suggest that these enhancer SNPs have undergone positive selection in the human lineage. Analysis of transcription factor binding sites created or abolished by the enhancer SNPs in this region using JASPAR (Sandelin et al., 2004; Vlieghe et al., 2006) (http://jaspar.cgb.ki.se/) suggests rs905740 is a leading functional candidate: the variant “T” allele increases the likelihood of Ets-1
binding (Figure 8), which has been shown to interact with the transcriptional coactivator EA1 binding protein p300 to regulate transcription at distal enhancer sites (Kumar et al., 2009; Yang et al., 1998). Significant upregulation of both Ets-1 and CHRNA5 in lung cancer (Bolon et al., 1995; Yamaguchi et al., 2007; Falvella et al., 2009) provides a disease model where the relationship between enhancer SNP rs905740 genotype and Ets-1 expression can be tested.

Figure 8. Ets-1 binding site created by rs905740, predicted by JASPAR
Approximately 50% of tissue-specific distal enhancers are found within 20kb of the nearest transcription start site, but can also occur >90kb up- or downstream of the genes they regulate (Yang et al., 1998). The pattern of multiple transcription factor binding sites around a given gene is proposed to determine tissue-selective enhancers (Pennacchio et al., 2007). Therefore, we tested the effect of rs905740 genotype on CHRNA5 mRNA expression in human lymphocytes, where it is also expressed. Lack of association between rs905740 genotype and expression in lymphocytes is consistent with this region serving as a tissue-specific distal enhancer. The activity of the enhancer in other tissues remains to be determined, but cortical and subcortical structures of the forebrain appear to show variable allelic expression (unpublished observations). It will be important to determine whether the proposed enhancer region regulates CHRNA5 expression in brain regions crucial for addiction, or whether CHRNA5 expression in lung tissues associates with lung cancer independent of nicotine addiction.

**Association of the enhancer region with clinical traits.** Previous clinical association studies have failed to reveal a role for the enhancer region alone in nicotine addiction, when using marker SNP rs880395 (Bierut et al., 2008; Saccone et al., 2009a and 2009b). However, the results in the present study strongly implicate rs880395, and those adjacent SNPs in complete LD, as strong drivers of CHRNA5 expression. This prompted reevaluation of rs880395 in nicotine dependence in conjunction with other functional SNPs. A joint analysis
of rs880395 and rs16969968 revealed a modest, but significant association for rs880395 with nicotine dependence. This is similar to a previous analysis examining the promoter variant rs3841324 (22bp deletion) with rs16969968 (Wang et al., 2009b). Both studies confirm that the minor allele of the nonsynonymous variant rs16969968 (A/A genotype) is a risk factor when occurring on the low expressing background, i.e. the ancestral allele for the proposed enhancer SNPs. SNP rs880395 and those in high LD with SNPs in this proposed upstream enhancer region emerged as another independent risk factor.

It may, at first glance, be surprising that the joint rs880395/ rs16969968 analysis did not yield stronger associations with nicotine addiction compared to a previous analysis examining the rs3841324/ rs16969968 diplotype (Wang et al., 2009b), since rs3841324 does not account for allelic expression differences. One explanation is that rs880395 and rs3841324 genotypes are highly correlated in populations of European descent (Table 1). Examining ethnic groups where the correlation between these two variants is not as strong might magnify the difference in risk potential for nicotine dependence. Given the large effect of the enhancer region on mRNA expression, future studies should consider the haplotype on which the risk polymorphism resides as an important factor in penetrance.

Our data cannot rule out the possibility that alternative splicing contributes to allelic differences in specific transcript expression, as indicated by differences
in AEI ratios between the two marker SNPs, even while the ratios correlate 
($r=0.62$). Human mRNA clones and expressed sequence tags display evidence 
for alternative splicing of $CHRNA5$ in exon 5, where rs16969968 resides. 
Presumably, any genetic variant driving a decrease in inclusion of rs16969968 in 
the mature mRNA could also affect risk for dependence.

Conclusions. $CHRNA5$ exists in three similarly abundant main haplotypes 
(Table 4) that have distinct biological functions, defined by the distal enhancer 
SNPs determining expression, and the nonsynonymous rs16969968 SNP 
determining ligand-mediated signaling: the ancestral haplotype with low 
expression, a high-expressing haplotype, and a low-expressing haplotype with 
altered protein and channel activity. The effects of high versus low $CHRNA5$ 
expression on channel activity in vivo remains to be determined. Because 
linkage disequilibrium at the $CHRNA5$ locus varies by ethnicity, the proposed 
enhancer variants should be considered when evaluating the penetrance of 
$CHRNA5$ in nicotinic receptor-related disorders. The presence of additional 
functional variants cannot be excluded, possibly resulting in a more complex 
haplotype repertoire with distinct functions.
Chapter 3: Variability in Serotonin 2A Receptor Genetics and Therapeutic Response

3.1 Introduction and Background

The G-protein coupled serotonin 2A receptor (5-HT2A) has fundamental roles in human cognition (Harvey, 2003; Meneses, 2007; Roth et al., 2004; Wingen et al., 2007), and is also linked to human neuropsychiatric disorders by the behavioral profile exhibited through direct agonism of the receptor. Psychotropic compounds, including lysergic acid diethylamide (LSD) and 4-iodo-2,5-dimethoxyphenylisopropylamine (DOI), act directly at the 5-HT2A receptor to alter perception and produce hallucinations (Sadzot et al., 1989; Titeler et al., 1988). However, not all 5-HT2A agonists produce this effect, calling into question the mechanism underlying this relationship. Initially these receptors were associated with G_q intracellular signaling pathways, which activates phospholipase C, ultimately resulting in release of intracellular calcium stores into the cytoplasm and activation of protein kinase C (Hoyer et al., 2002). More recently, researchers demonstrated functional selectivity of 5-HT2A receptors, whereby they also signal via the G_i pathway in a ligand-dependent manner to produce their hallucinogenic effects (González-Maeso et al., 2007).
In the human brain, 5-HT2A receptors are most abundant on pyramidal cells in layer V of the cortex (Jakab & Goldman-Rakic, 1998; Pasqualetti et al., 1996; Pompeiano et al., 1994; Xu & Pandey, 2000), where there is evidence that they physically and functionally interact with glutamate and dopamine receptors (González-Maeso et al., 2008; Lukasiewicz et al., 2010) to modulate cortical output from layer V and corresponding behaviors. As it relates to psychotic symptomatology, physical interactions with metabotropic glutamate 2 receptors mediates the extent to which 5-HT2A receptors signal via the G\textsubscript{i} pathway to induce hallucinations. Thus, the 5-HT2A receptor is at the crux of antipsychotic pharmacotherapeutics and serves as a primary target of atypical antipsychotics, even to a greater extent than the dopamine D2 receptors more heavily targeted by first-generation antipsychotics (Altar et al., 1986; Schotte et al., 1996; Stockmeier et al., 1993).

The 5-HT2A receptor is also crucial for antidepressant therapies. Acute and direct antagonism of 5-HT2A receptors can alleviate depressive-like behaviors (Pandey et al., 2010), whereas receptor agonism can produce depressive-like behaviors (Rajkumar et al., 2009) in rodent models. Selective serotonin reuptake inhibitors (SSRIs), which constitute the largest class of antidepressants, increase synaptic concentrations of serotonin (5-HT), prolonging signaling at postsynaptic 5-HT2A receptors (Boothman et al., 2006; Meyer et al., 2001). While acute agonism findings are contrary to the presumed environment produced by SSRIs (i.e. beneficial versus detrimental effects of
increased agonism), receptor expression and differential second messenger signaling in normal versus depressed individuals could explain this disparity (Shelton et al., 2009). Nonetheless, 5-HT2A receptors are an integral component of depressive symptomatology.

Interindividual responses to antipsychotics and antidepressants hamper their clinical efficacy in a large percentage of the population. For example, approximately 50% of patients do not respond to risperidone for the treatment of schizophrenia in placebo-controlled double-blind clinical trials (Marder & Meibach, 1994). A similarly high percentage of patients fail to respond to different SSRIs (Claghorn et al., 1992; Hirschfeld, 1999). Even when patients do respond to treatment, unwanted side-effects and adverse drug reactions can significantly reduce continuation of the treatment. To increase quality of care and reduce unwanted risks associated with treatment, clinicians are moving towards personalized medicine, whereby drug efficacy or the likelihood of adverse drug reactions can be predicted for individual patients. Not surprisingly, genetics are a major predictor of these drug-related responses.

The gene encoding 5-HT2A (HTR2A) has been extensively studied in a pharmacogenomics context. Numerous studies indicate that genetic variability within the HTR2A locus contributes heavily to individual responses to atypical antipsychotics and antidepressants (Lane et al., 2005; Malhotra et al., 2004; Serretti & Artioli, 2004). HTR2A harbors one frequent synonymous polymorphism, rs6313 (also referred to as T102C), and several nonsynonymous
polymorphisms, only one of which (rs6314) has an allele frequency greater than 1% in populations of European descent. Figure 9 maps these SNP locations, and additional HTR2A SNPs identified as functional pharmacogenomic candidates, including the promoter variant rs6311 and intronic variants rs7997012 and rs17288723. The synonymous SNP rs6313, in particular, is associated with inherited risk for schizophrenia (Golimbet et al., 2007; Vaquero Lorenzo et al., 2006), response to atypical antipsychotics in schizophrenic patients (Arranz et al., 1998; Chen et al., 2009; Kim et al., 2008; Lane et al., 2002; Yamanouchi et al., 2003), and side-effects of antipsychotic treatment (Gunes et al., 2007; Melkersson et al., 2010; Ujike et al., 2008), although these findings are sometimes not replicable (Al Hadithy et al., 2008; Ikeda et al., 2008; Ji et al., 2008; Pae et al., 2005). Similar results are observed for major depression and clinical response to antidepressants (Kato & Serretti, 2008).

Figure 9. HTR2A gene map denoting pharmacogenomic candidate SNPs

Few studies have investigated functional mechanisms related to the pharmacogenomic candidate SNPs in HTR2A. Two previous studies measuring allelic expression of HTR2A using rs6313 as a marker SNP both concluded that
there are no frequent functional variants modulating the overall expression of
HTR2A mRNA (Bray et al., 2004; Fukuda et al., 2006). This conclusion is
particularly important, as it suggests that the promoter variant rs6311 is not
functionally related to transcription, despite contrary in vitro evidence (Myers et
al., 2007). Perhaps the best functional evidence exists for the nonsynonymous
SNP rs6314 (His452Tyr), whereby the variant allele displays reduced intracellular
second messenger activity (Hazelwood & Sanders-Bushaffinity, 2004; Ozaki et
al., 1997) and reduced affinity for atypical antipsychotics (Davies et al., 2006 and
2010). A relationship between rs6314 and clinical response to antipsychotics
and antidepressants, however, is inconsistent (Masellis et al., 1998; Minov et al.,
2001; Peters et al., 2004; Wilkie et al., 2009). Functional evidence for the other
pharmacogenomic candidate SNPs rs6313, rs7997012, and rs17288723 is
lacking.

For any single SNP, the evidence between the HTR2A locus and a clinical
or functional relationship is equivocal, but the composite body of evidence
suggests this region harbors one or more functionally important variants that are
predicted to be relatively common in the general population. Here, we surveyed
a large cohort of post-mortem brain tissue samples for interindividual variability in
HTR2A mRNA expression that correlates with genetic variants. Similar to
previous studies, we find no major genetic factors driving overall expression of
HTR2A. However, we do find evidence of a novel spliceoform that encodes a
truncated 5-HT2A receptor, the expression of which correlates with the
synonymous pharmacogenomic candidate SNP rs6313. The alternative splice site for this isoform is a highly conserved region of *HTR2A* corresponding to translation start sites in lower vertebrates. In light of this finding, we reassess the functional role of rs6313 and its relationship to pharmacological interventions.

### 3.2 Materials and Methods

**Nucleic acid isolation.** Genomic DNA (gDNA) was isolated from 91 prefrontal cortex tissues (provided by Dr. Deborah Mash) originating from Brodmann Area 46, using a “salting out” method (Miller et al., 1988) supplemented with additional sodium dodecyl sulfate for lipid-rich brain tissue. Total RNA was isolated with RNeasy Mini Kit spin columns (Qiagen, Germantown, MD) from brain homogenates dissolved in TRizol (Life Technologies, Carlsbad, CA). RNA preparations were treated with DNase to remove any remaining gDNA fragments. Two separate cDNA preparations were made using 0.5µg total RNA per sample preparation. Gene-specific primers targeting the *HTR2A* locus (Appendix B: Table 11) and oligo-dT was used to prime mRNA in the reverse transcription reaction.

**Brain sample genotyping.** SNPs rs6313 and rs6314, used as markers for measuring AEI, as well as intronic SNPs rs655888 and rs2760351 were genotyped using SNaPshot, as described in section 2.2 and Lim et al. (2007). To genotype rs1328684, the region surrounding the SNP was PCR-amplified using primers tagged with a fluorophore (6-FAM or HEX) and HinfI was used to cut the
wild-type “G” allele. Fragment length, used to determine genotype, was resolved by an ABI 3730 DNA Analyzer (Life Technologies). Allele-specific fluorescence-tagged primers were used to genotype AEI marker SNP rs3803189 and intronic SNP rs7330461.

Allelic mRNA Expression Imbalance in human brain tissue. Allele-specific mRNA expression was quantified in samples heterozygous for any marker SNP (rs6313, rs6314, or rs3803189) using a primer extension assay (SNaPshot), as described in section 2.2. Primer sequences for quantitative measures (AEI and qPCR) are listed in Appendix B: Table 11. Following the SNaPshot primer extension step and separation on an ABI 3730, area under the curve for each allele was determined using GeneMapper 4.0 software (Life Technologies) to calculate relative allelic expression ratios. For each SNP, samples were measured in triplicate for each of two independent cDNA syntheses, while gDNA was measured in duplicate. Allelic ratios for cDNA were normalized against the overall average ratio obtained for gDNA for each marker SNP (ancestral/variant allele).

Relative Spliceoform quantification. A forward primer targeting exon 1 and fluorescence-tagged reverse primer targeting exon 3 of the HTR2A transcript was used to PCR-amplify cDNA and subsequently detect alternative splicing of exon 2 in all brain samples. PCR amplicon length was resolved using an ABI 3730. Peak height ($P_n$) was used as an indication of spliceoform expression,
measured in triplicate, and relative expression was calculated within each assay as individual peak height divided by the sum of all peak heights ($P_n/\Sigma P_1…P_x$).

**Sequencing.** cDNA (25ng) was amplified with Sigma Ready-Mix (Sigma-Aldrich, St. Louis, MO) using the same primers used in the relative spliceoform quantification analysis, targeting HTR2A exons 1-3, but absent the fluorophore. Specific products corresponding to each spliceoform were excised following resolution on a 2% agarose gel, purified using a Qiaquick Gel Extraction Kit (Qiagen) or GenElute Gel Extraction Kit (Sigma-Aldrich, St. Louis, MO), and then sequenced by the Ohio State University Plant-Microbe Genomics Facility using an ABI 3730 DNA Analyzer (Life Technologies). Sequence data were aligned using the Basic Local Alignment Search Tool – bl2seq (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and compared against Reference Sequence (NCBI Build 37 on the UCSC Genome Browser (Kent et al., 2002): http://genome.ucsc.edu).

**Quantitative real-time PCR (qPCR).** Total HTR2A mRNA expression was measured in triplicate via qPCR on an ABI 7000 Sequence Detection System (Life Technologies). 12.5ng of cDNA was amplified in a 15µl reaction using Power SYBR Green PCR Master Mix (Life Technologies). Primers for total HTR2A measurements targeted the same region as used for SNaPshot analysis of rs6314. Primers specific for individual splice variants spanned exon-exon boundaries where alternative splicing occurs, maximizing the likelihood that only a single splice variant is being amplified in each reaction. The relative quantity of
HTR2A mRNA, normalized within each sample to β-actin (ACTB) expression (measured in duplicate), was used for data analysis.

Data Analysis. HelixTree software (Golden Helix, Bozeman, MT) was used to test genotype associations with both relative and absolute measures of mRNA expression and to perform linear regression analysis of genotype on measures of HTR2A mRNA expression. Univariate analysis of variance comparing HTR2A mRNA expression across demographic categories, with age and post-mortem interval as covariates, and subsequent correlations and one-way ANOVAs comparing HTR2A expression across genotype for brain samples was performed with SPSS v17.0 (SPSS, Inc., Chicago, IL).

3.3 Results

Allelic Expression Imbalance in Human Prefrontal Cortex. Demographics for the brain samples used in the study are included in Table 5. Analysis of variance (ANOVA) found no significant main effect of age, race, gender, cocaine use, smoking history, or post-mortem interval on total HTR2A mRNA expression in these samples.
<table>
<thead>
<tr>
<th>Demographic</th>
<th>Group (n)</th>
<th>HTR2A Expression (CT Avg. ± St. Dev.)</th>
<th>ANOVA (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male (62)</td>
<td>6.8 ± 2.3</td>
<td>0.665</td>
</tr>
<tr>
<td></td>
<td>Female (29)</td>
<td>6.9 ± 1.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Caucasian (59)</td>
<td>6.8 ± 1.8</td>
<td>0.167</td>
</tr>
<tr>
<td>Race</td>
<td>African American (14)</td>
<td>7.2 ± 2.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mixed/Other (18)</td>
<td>6.8 ± 2.9</td>
<td></td>
</tr>
<tr>
<td>Drug Group</td>
<td>Cocaine (39)</td>
<td>7.2 ± 2.6</td>
<td>0.470</td>
</tr>
<tr>
<td></td>
<td>Control (52)</td>
<td>6.6 ± 1.5</td>
<td></td>
</tr>
<tr>
<td>Smoking History</td>
<td>Smoker (35)</td>
<td>6.8 ± 2.4</td>
<td>0.576</td>
</tr>
<tr>
<td></td>
<td>Non-smoker (56)</td>
<td>6.9 ± 1.9</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>0.806</td>
</tr>
<tr>
<td>Postmortem Interval</td>
<td></td>
<td></td>
<td>0.228</td>
</tr>
</tbody>
</table>

Table 5. Influence of demographic variables on HTR2A mRNA expression in BA46 prefrontal cortex samples

Fifty-three of 91 samples were heterozygous for at least one of the marker SNPs (rs6313, rs6314, or rs3803189), permitting AEI analysis. Overall, no sample displayed differential allelic expression greater than 1.5-fold (range: 0.80 – 1.47) for any of the marker SNPs, suggesting a lack of genetic variants with strong effects on overall HTR2A expression. Variation of allelic ratios measured in DNA for all marker SNPs ranged from 10.5 – 12.9% across samples, and within-sample standard deviation for AEI measurements in cDNA ranged from 4.8 – 6.9%, indicating high assay replicability. Thirty-six samples were heterozygous for rs6313, 9 samples for rs6314, and 29 samples for rs3803189, with normalized AEI measurements ranging from 0.81 – 1.47, 0.87 – 1.18, and 0.80 – 1.40 for each SNP, respectively (Figure 10). When comparing samples heterozygous for more than one marker SNP (Figure 11), differences in the
magnitude of allelic expression are evident, suggesting the presence of allelic alternative splicing of HTR2A mRNAs.

Figure 10. AEI at three HTR2A marker SNPs; a) rs6313, b) rs6314, and c) rs3803189. Note: Dotted lines represent a 1.5-fold difference in allelic expression.
Figure 10 continued

**Figure 11.** AEI comparison in samples heterozygous for multiple *HTR2A* marker SNPs. Arrows denote samples where the relative magnitude of AEI differs between marker SNPs, suggesting the occurrence of allelic splicing. *Note:* dotted lines represent a 1.5-fold difference in allelic expression.
**Spliceoform Identification and Relative Quantification.** PCR amplification of HTR2A mRNA targeting exons 1-3 revealed at least three spliceoforms that were present in all samples. The genomic sizes for two of the spliceoforms corresponded to previously known HTR2A transcripts, in which exon 2 is present or spliced out (E2^+, E2^-), while the size of the third spliceoform did not correspond to any previously known splice variant. Sequencing of this third variant confirmed that it originated from the HTR2A gene, but was alternatively spliced, producing a truncated exon 2 (E2^tr). Figure 12 displays alternative splicing of HTR2A transcripts E2^+, E2^-, and E2^tr.

![Diagram of HTR2A spliceoforms](image)

Figure 12. Alternative splicing of HTR2A transcripts. Red exons indicate protein-coding regions, while black exons are untranslated.

The average percentage of relative expression for each spliceoform in this prefrontal cortex tissue was 58% (E2^+), 24% (E2^tr), and 18% (E2^-), with the majority of samples following the general pattern E2^+ > E2^tr > E2^- . One sample significantly differed from this pattern, with E2^- transcript constituting 85% of the relative spliceoform expression. Sequencing gDNA from this patient revealed
two unannotated SNPs located less than 1kb upstream of exon 3, perhaps responsible for directing alternative splicing in this region. A 1.68-fold difference in allelic expression for the E2\textsuperscript{+} transcript (probed for SNP rs6312 in exon 1) was observed in this patient, supporting allelic splicing.

Genotype Associations to HTR2A Expression. The relative amount of each spliceoform was compared against genotypes for rs6313, rs1328684, rs655888, rs7330461, rs6314, and rs3803189 using linear regression, revealing a significant relationship between E2\textsuperscript{+} expression and rs6313 genotype (p=0.004). Subsequent univariate ANOVA confirmed a significant main effects of rs6313 genotype (p=0.032, F=3.81), and also revealed significant main effects of race (p=0.018, F=4.52), age (p=0.012, F=7.15), and post-mortem interval (p=0.028, F=5.30) on expression of E2\textsuperscript{+}. Patients homozygous for the rs6313 minor “T” allele (T/T) have significantly higher relative expression of E2\textsuperscript{+} versus homozygous C/C patients, with heterozygotes (C/T) displaying intermediate expression (C/C=21% vs. C/T=24% vs. T/T=29%; p=0.032, F=3.81; Figure 13). Relative E2\textsuperscript{+} expression was also significantly lower in African-Americans compared to European-Americans (16% vs. 26% of total; p=0.008, F=5.539).

Linear regression analysis comparing total HTR2A mRNA expression, measured via qPCR, to rs6313, rs1328684, rs655888, rs7330461, rs6314, and rs3803189 genotypes revealed a significant relationship between rs1328684 and expression (p=0.048). However, subsequent univariate ANOVA in a larger sample failed to support this relationship (p=0.183). Also using linear regression,
no other significant expression X genotype associations were found for any specific spliceoform.

![Bar chart showing relative amount of E2^tr spliceoform compared across rs6313 genotype.

Figure 13. Relative amount of E2^tr spliceoform compared across rs6313 genotype. Note: * = p<0.05.

3.4 Discussion

Aberrant serotonergic signaling is recognized as a major component of psychiatric illness, driving efforts to design and manufacture pharmaceuticals targeting this neurotransmitter system. In 2009, antipsychotics represented the largest market share within the pharmaceutical industry, totaling $14.6 billion,
while antidepressants ranked fourth at $9.9 billion (IMS Health Press Release, 2010). As we enter the era of personalized medicine, the National Institutes of Health is spearheading a $161.3 million effort to understand drug response from a pharmacogenomic perspective (National Institute of General Medical Sciences Press Release, 2010). Understanding HTR2A genetic variability that underlies interindividual responses to pharmacological treatment remains a major goal of the pharmacogenomics field. Here, we demonstrate a link between a pharmacogenomic candidate SNP (rs6313) and expression of a specific HTR2A mRNA spliceoform, perhaps the first such study to link a molecular genetic mechanism to this SNP.

The truncated exon 2 is unlikely to be a spurious result of noisy transcription. A closer look of the DNA sequence at the alternative splice site reveals the presence of all features necessary for alternative splicing to occur, including splice acceptor and donator sites, a polypyrimidine tract, and a consensus branch point (Figure 14). The mechanism through which a point mutation at rs6313 increases alternative splicing in favor of the truncated variant is not clear, as rs6313 resides approximately 125 bases upstream of the alternatively-splice exon 2, beyond the branch point and polypyrimidine tract. One possibility is a change in local RNA folding, which could mediate access of alternative splicing proteins to motifs in the RNA. Conformational analysis of exon 2 by mfold (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi; Mathews et al., 1999; Zucker, 2003) reveals differential local interaction between the rs6313
alleles and the region where exon 2 is alternatively spliced (Figure 15),
supporting RNA folding as a putative mechanism to explain the change in
alternative splicing.

Figure 14. Consensus sites for alternative splicing of HTR2A E2tr transcript

Figure 15. E2tr and rs6313 interaction predicted by HTR2A exon 2 folding

Another possibility includes the formation or abolition of a consensus sequence
for a splicing protein by the variant allele. Prediction of splicing protein binding
sites by the freely available database SpliceAid (http://www.introni.it/splicing.html; Piva et al, 2009) suggests the minor “T” allele for rs6313 deletes consensus binding sites for three serine/arginine-rich splicing factors (SRSF); SRSF1, SRSF6, and SRSF9, while it creates a consensus binding site for SRSF5, all of which are expressed to some degree in the human prefrontal cortex.

Sequence analysis of this truncated mRNA transcript suggests it is likely translated into a protein in vivo. A phylogenetic analysis of HTR2A reveals that the translation start site for the receptor protein in lower vertebrates corresponds to the exact base where alternative splicing occurs (Appendix C: Figure 23). The putative translation start site for the human full-length protein is 216 nucleotides upstream of this site and is no longer present in this shorter splice variant, but translation could still be initiated at an in-frame site downstream of the conserved transcription start site. Analysis of the RNA sequence at this downstream start site reveals the presence of a strong Kozak consensus sequence necessary for initiation of protein translation. The predicted protein is considerably truncated at the N-terminus, lacking the first 96 amino acids compared to the full-length protein. This truncation specifically deletes the entire extracellular N-terminus, most of the first transmembrane domain, the first 6 amino acids of the 7-transmembrane domain conserved among GPCRs, and all putative N-linked glycosylation sties (Figure 16). Deleting these domains is likely to have functional consequences. A recent in vitro study examining the 5-HT2A receptor demonstrated that blocking glycosylation, either by mutating the protein or
through chemical inhibition, significantly reduces receptor membrane expression (Maginnis et al., 2010). Alone, this effect might be inconsequential, because the majority of HTR2A transcript present in the human cortex encodes the full-length protein. However, there is evidence that the 5-HT2A receptor homodimerizes (Brea et al., 2009) and heterodimerizes with the metabotropic glutamate receptor 2 (mGluR2) (González-Maeso et al., 2008) and the dopamine D2 receptor (DRD2) (Lukasiewicz et al., 2010), perhaps allowing the truncated protein to act in a dominant negative fashion against the full-length proteins. These complexes, between 5-HT2A and mGluR2 or DRD2, can display signaling cascades functionally distinct from their canonical pathways (González-Maeso et al., 2008; Lukasiewicz et al., 2010) and are proposed targets for antipsychotic drugs.

Figure 16. 5-HT2A protein and predicted E2tr truncation. Areas shaded in red are predicted to be deleted from the full-length protein, including N-linked glycosylation sites at amino acids 8, 38, 44, 51, and 54.
Similar to previous reports examining \textit{HTR2A} (Bray et al., 2004; Fukuda et al., 2006), no genetic variant was found to have a major effect on total mRNA expression. Only for three samples did we observe an allelic expression difference greater than 1.3-fold. It is still quite possible that this modest difference in expression has a dramatic effect on phenotype, but the dearth of samples displaying AEI did not allow us to link expression to a specific allele or genotype that we could subsequently test in clinical populations. Given the transcript heterogeneity due to alternative splicing, it is still possible that a functional SNP in this region is driving allelic expression of any one spliceoform. Low expression and/or lack of marker SNPs in specific spliceoforms restricted allelic analyses in our study. In-depth sequencing of the transcriptome using next-generation technologies can facilitate future allelic analyses of \textit{HTR2A}, allowing more precise attribution of genetic effects contributing to spliceoform-specific expression.

This study delineates a mechanistic relationship between alternative splicing of \textit{HTR2A} and genetic variability at the pharmacogenomic candidate SNP rs6313. Furthermore, this study highlights the potential consequence of a synonymous SNP on RNA processing – one that is assumed to be under more heavy selective pressure than an intronic SNP, since it resides in a coding region. In all HapMap populations, C/T is the predominate genotype, constituting approximately 40-50\% of all populations, with minor allele frequencies ranging from 35-50\%. The observed high allele frequencies in all populations might
suggest positive selection for the minor “T” allele, from which one could speculate that increased formation of the truncated spliceoform confers a reproductive advantage. However, the relationship between rs6313 and HTR2A mRNA splicing remains to be replicated and a functional role of the truncated protein remains to be demonstrated in vivo and in vitro. Nonetheless, the 5-HT2A receptor remains a critical drug target for treating neuropsychiatric disorders. New insight into the biological relationships between receptor complexes are providing new and exciting avenues for treatment, but a firm understanding of interindividual variability in HTR2A genetics is necessary to optimize new and existing treatment.
Chapter 4: Surveying Candidate Risk Genes in Autism Spectrum Disorders

4.1 Introduction and Background

Genome-wide association studies, analyses of copy number variants, DNA sequencing approaches, and familial linkage studies have revealed a number of autism risk genes encoding cellular adhesion molecules (CAMs) (Betancur et al., 2009). These include the neurexin, neuroligin, contactin, and cadherin gene families, in addition to members of the general immunoglobulin CAM superfamily. Subsequent attempts to find common SNPs conferring risk, by sequencing DNA collected from affected individuals, have revealed few disease-causing coding variants (Blasi et al., 2006; Durand et al., 2007; Feng et al., 2006; Gauthier et al., 2005; Gauthier et al., 2009; Vincent et al., 2004; Yan et al., 2005; Yan et al., 2008; Ylisaukko-oja et al., 2005). These sequencing studies focused exclusively on protein coding regions and adjacent constitutive splice sites, driven by the assumption that the pathogenic mutations are deleterious amino-acid substitutions in these CAM proteins. A lack of disease-causing mutations uncovered via sequencing led researchers to suggest ASD results from highly penetrant, rare mutations, rather than those more common but less penetrant (Zhao et al., 2007). More recently, however, increased understanding of RNA biology continues to reveal mechanisms by which RNA processing
defects cause or contribute to disease phenotypes in humans, especially in neurodegenerative and neuromuscular disorders, where dysfunctional RNA splicing is prominent (Cooper et al., 2009).

Under normal conditions, constitutive and alternative splicing is directed by a nuclear protein aggregate of more than 100 core proteins and secondary modulators, collectively known as the spliceosome, which interacts with heterogeneous nuclear RNA (hnRNA or pre-mRNA) in a process highly conserved in eukaryotes (Zhou et al., 2002). This process produces tissue-specific RNA and protein isoforms that can vary throughout development or in response to synaptic activity (Li et al, 2007). Spliceosomal proteins recognize consensus cis-regulatory motifs, located in introns or exons, and subsequently govern the inclusion or exclusion of alternative exons in the mature mRNA end-product. Pathogenic mutations often disrupt regulatory cis-binding splice motifs within known risk genes, resulting in aberrant splicing, and manifestation of the disease phenotype. Genetic, epigenetic, or environmental factors can also affect the capacity of trans-acting splice factors and produce aberrant splicing patterns contributing to disease. Figure 16 portrays an example of the relationship between trans- and cis-acting factors during alternative splicing. Global estimates based on mathematical models suggest up to 60% of all disease-causing mutations disrupt splicing in cis or trans (Lopez-Bigas et al., 2005; Wang and Cooper, 2007). When individual genes are assayed for disease-causing mutations, up to 50% of pathogenic mutations affect splicing, largely through cis-
acting mechanisms. (Ars et al., 2000; Cartegni et al., 2002; Mayer et al., 1999; Mayer et al., 2000; Messiaen et al., 2000; Pagenstecher et al., 2006; Teraoka et al., 1999).

Figure 17. Effect of mutations in cis- and trans-acting factors on RNA splicing. Mutations in trans-acting splice factors change splicing activity in a pleiotropic manner (middle panel), while cis-acting mutations affect only the mutated gene (right panel).

As it relates to autism, mRNA transcribed from CAM genes is extensively spliced to produce unique protein end-products necessary for specialized cell-to-cell interactions (Shapiro et al., 2007). In the central nervous system, alternative splicing of CAMs is associated with maturation of synapses and synaptic fidelity. For example, trans-synaptic binding of the CAMs α-neurexin and neuroligin is determined by alternative splicing of the acetylcholinesterase (AchE)-like domain in neuroligin mRNA transcripts, which encodes a critical N-glycosylation site that blocks α-neurexin binding (Boucard et al., 2005). The different combinations of
trans-synaptic binding subsequently determines the neurotransmitter phenotype at the synapse, either towards excitatory glutamatergic or inhibitory gamma-aminobutyric acid (GABA)ergic signaling (Chih et al., 2006).

One likely functional consequence of CAM mutations on a synaptic level, based on animal and in vitro human studies, is an imbalance in homeostatic glutamatergic and GABAergic signaling (Blundell et al., 2009; Chih et al., 2005; Chubykin et al., 2007; Gibson et al., 2009; Hines et al., 2008; Huang and Scheiffele, 2008; Tabuchi et al., 2007). Instability of homeostatic glutamatergic versus GABAergic signaling has been put forth as one hypothesis to explain, at least in part, the etiology of autism (Rubenstein & Merzenich, 2003). One promising candidate gene that regulates glutamatergic signaling, the neuronal glutamate transporter SLC1A1, is significantly associated with obsessive-compulsive behaviors (Arnold et al., 2006; Dickel et al., 2006; Stewart et al., 2007) and ASD (Brune et al., 2008; Gadow et al., 2010; Kantojärvi et al., 2010). Genetic variants within the glutamate/aspartate transporter SLC25A12 are also associated with ASD (Ramoz et al., 2004; Segurado et al., 2005; Turunen et al., 2008), although some studies fail to replicate this relationship (Chien et al., 2010; Correia et al., 2006; Rabionet et al., 2006). Protein and mRNA analyses from affected individuals also provide evidence for disrupted glutamatergic and GABAergic signaling in autism (Blatt et al., 2001; Purcell et al., 2001).

Microarray studies comparing gene expression between ASD and controls cohorts have highlighted gene splicing and cellular adhesion as biological
processes contributing to ASD (Abrahams and Geschwind, 2008). However, no study has systematically examined ASD risk genes for the presence of functional SNPs that affect mRNA expression, especially in tissues germane to the disorder. To examine risk genes for the presence of common functional variants contributing to ASD, we measured allelic expression of 19 ASD risk genes from three functional categories highlighted above, which includes cellular adhesion, alternative splicing, and glutamatergic and GABAergic signaling (Appendix B: Table 12). Our results indicate the presence of functional SNPs that significantly affects mRNA expression in five ASD risk genes, including the neuronal glutamate transporter gene SLC1A1, the cellular adhesion molecule contactin associated protein CNTNAP2, and three splicing related proteins, heterogeneous nuclear ribonucleoprotein A3 (HNRNPA3), neuro-oncological ventral antigen 1 (NOVA1), and splicing factor 4 (SF4). For the remaining genes that did not display allelic expression differences, we can begin ruling out the possibility that common functional variants significantly contribute to their overall expression.

4.2 Materials and Methods

Nucleic acid isolation. Genomic DNA (gDNA) was isolated from 56 frontopolar cortex tissues (provided by The Autism Tissue Program; http://www.autismtissueprogram.org) originating from Brodmann Area 10, using a “salting out” method (Miller et al., 1988) supplemented with additional sodium dodecyl sulfate for lipid-rich brain tissue. Demographics for these samples are
listed in Table 6. Total RNA was isolated with RNeasy Mini Kit spin columns
(Qiagen, Germantown, MD) from brain homogenates dissolved in TRIzol (Life
Technologies, Carlsbad, CA). RNA preparations were treated with DNase to
remove any latent gDNA fragments. Complimentary DNA (cDNA) preparations
were made using 0.5µg total RNA per sample preparation. Gene-specific
primers and oligo-dT was used to prime mRNA in the reverse transcription
reaction.

**Brain sample genotyping.** All SNPs in this study were genotyped using
SNaPshot, as previously described in section 2.2 and Lim et al. (2007), with
primers used for AEI analysis listed in Appendix B: Table 12. Genotypes were
resolved by an ABI 3730 DNA Analyzer (Life Technologies).

**Quantitative real-time PCR (qPCR).** mRNA expression for each gene was
measured via qPCR on an ABI 7000 Sequence Detection System (Life
Technologies). 12.5ng of cDNA was amplified in a 15µl reaction using Power
SYBR Green PCR Master Mix (Life Technologies). Cycle threshold (C\(_T\)) values
were normalized to the housekeeping gene β-actin (ACTB) for analysis. Primers
for qPCR measurements targeted the same region as used for SNaPshot
analysis for each gene.
<table>
<thead>
<tr>
<th>Cohort</th>
<th>Sex</th>
<th>Descent</th>
<th>Age (Avg. ± St. Dev.)</th>
<th>PMI (Avg. ± St. Dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autism</td>
<td>M = 15</td>
<td>African = 3</td>
<td>26.5 ± 21.0</td>
<td>16.9 ± 5.9</td>
</tr>
<tr>
<td></td>
<td>F = 6</td>
<td>European = 18</td>
<td>Range: 2 - 82</td>
<td></td>
</tr>
<tr>
<td>Relative</td>
<td>M = 7</td>
<td>African = 0</td>
<td>64.5 ± 24.9</td>
<td>16.8 ± 4.0</td>
</tr>
<tr>
<td></td>
<td>F = 8</td>
<td>European = 15</td>
<td>Range: 9 - 90</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>M = 18</td>
<td>African = 1</td>
<td>42.0 ± 23.5</td>
<td>21.4 ± 5.0</td>
</tr>
<tr>
<td></td>
<td>F = 2</td>
<td>European = 19</td>
<td>Range: 7 - 88</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Brodmann Area 10 cortex tissue demographics

*Allelic mRNA Expression Imbalance in human brain tissue.* Allele-specific mRNA expression was quantified in samples heterozygous for marker SNPs listed in Appendix B: Table 12 using a primer extension assay (SNaPshot), as described in section 2.2. Following the SNaPshot primer extension step and separation on an ABI 3730, peak height and area under the curve was determined using GeneMapper 4.0 software (Life Technologies) to calculate relative allelic expression ratios. For each SNP, samples were measured in duplicate for cDNA and gDNA. Allelic ratios for cDNA were normalized against the overall average ratio obtained for gDNA for each marker SNP (ancestral/variant allele).

*Data Analysis.* Samples with allelic expression ratios differing by more than 3 standard deviations from the within-sample error for cDNA or the between-sample error for gDNA for each assay were considered as significantly deviating from unity and likely harboring functional cis-acting polymorphisms.
4.3 Results

Total mRNA expression and allelic expression measured at each SNP location for each gene is displayed in Appendix B: Table 13. High cycle threshold values for CNTN4 and GABRB3, indicating low mRNA expression, did not allow subsequent AEI analysis for these genes. Multiple samples displayed significant AEI for SLC1A1, CNTNAP2, HNRNPA3, NOVA1, and SF4, suggesting the presence of a common cis-acting variant affecting mRNA expression of these genes. The remaining genes displayed relatively equal amounts of expression from each allele measured.

For SLC1A1, significant AEI was observed in 14 of 42 samples, measured at three different SNP locations (Figure 17). AEI for this gene appears bimodal, suggesting the presence of at least two different functional SNPs, each significantly contributing to allelic expression of SLC1A1. The first and more dramatic AEI phenotype represents a >1.6-fold difference in allelic expression, and is present in 3 of 42 samples, one each from the autism, control, and first degree relative cohorts. This more extreme AEI phenotype is replicable when measured for multiple SNPs in the same tissue, as shown in Figure 18. The second AEI phenotype is less dramatic, with 11 of 42 samples (3 autistic, 4 relatives, 4 controls) displaying significant AEI ranging between 1.15 and 1.41-fold differences between alleles. Within-sample variability, used to calculate cutoff scores for determining significant AEI, was well within an acceptable range for each SNaPshot assay (9% for rs301430, 8% for rs2228622, 3% for
rs3087879), allowing highly reproducible results. For any given SNP at which AEI was measured, significant AEI values corresponded to 3.0 – 17.5 standard deviations (S.D.) of the error from unity.

Figure 18. AEI in SLC1A1 measured at three marker SNP locations: a) rs301430, b) rs2228622, and c) rs3087879. Note: Dotted lines indicate ±3 S.D. of average within-sample error.
Figure 18 continued

SLC1A1 rs3087879 AEI (Height)

SLC1A1 Combined AEI (Height)

Figure 19. Within-sample comparison of AEI at multiple SLC1A1 SNP positions. Two samples had replicable significant AEI at multiple SNP locations (arrows). Note: Dotted lines indicate ±3 S.D. of average within-sample error.
For \textit{CNTNAP2}, significant AEI was observed for 4 of 25 samples heterozygous for rs2530312 samples (2 autistics, 1 relative, and 1 control) (Figure 19). Results were also highly reproducible for this assay, as within-sample variability measured \textasciitilde4\%, resulting in a cutoff for significant AEI of 1.11-fold differences in allelic expression. Samples with significant AEI displayed 1.16-, 1.21-, 1.22-, and 1.22-fold differences in expression between the alleles, equating to 4.3 – 5.9 S.D. of the error from unity.

![Figure 20. Significant AEI for \textit{CNTNAP2} in 4 samples (ASD013, ASD020, ASD033, and ASD039) measured at rs2530312. \textit{Note}: Dotted lines indicate \pm3 S.D. of average within-sample error.](image)

Four of 20 samples (1 autistic sample and 3 relatives) had significant AEI for \textit{HNRNPA3}, measured at marker SNP rs9068 (Figure 20). The presence of a pseudogene (\textit{HNRNPA3P1}) with significant homology to the region harboring
rs9068 was apparent in the AEI assay for gDNA, requiring mathematical correction (dividing the allelic gDNA ratio by 2 to account for the additional pair of HNRNPA3P1 alleles). After correction, between-sample error for gDNA was approximately 6%, producing a cutoff for significant AEI at 1.18-fold difference in allelic expression. For one sample, minor allele expression was greater relative to the major allele (AEI = 0.84; 1.19-fold allelic difference), while the opposite was observed for the remaining three samples (AEI = 1.28, 1.42, and 1.54). These AEI values were approximately 3.1, 4.8, 7.1, and 9.0 S.D. of the gDNA error from unity, respectively.

Figure 21. Significant AEI for HNRNPA3 in 5 samples (ASD009, ASD014, ASD023, and ASD034) measured at marker SNP rs9068. Note: Dotted lines indicate ±3 S.D. of average within-sample error.
Of 16 samples heterozygous for marker SNP rs1245213 in NOVA1, 2 autistic, 2 relatives, and 1 control sample had significant AEI, ranging from 1.35 – 1.45-fold differences in allelic expression (Figure 21). These AEI values were approximately 3.2 – 4.1 S.D. from unity. Generally, allelic differences favored greater expression of the major allele relative to the minor allele, with an additional three samples close to meeting criterion for significance (Figure 21). Within-sample error for this assay was approximately 11%.

![NOVA1 rs1245213 AEI (Height)](image)

Figure 22. Significant AEI for NOVA1, measured at marker SNP rs1245213. Note: Dotted lines indicate ±3 S.D. of average within-sample error.

Six of 11 samples (3 autistic, 2 relatives, and 1 control) had significant AEI for SF4, measured at marker SNP rs2010506 (Figure 22). Between-sample error for gDNA measurements was 12.7%, setting the cutoff for significant AEI above
1.38-fold differences in allelic expression. AEI deviated from unity in both the positive and negative direction (range: 0.46 – 1.68-fold allelic differences), indicating that the marker SNP is not likely to be in linkage disequilibrium with the functional variant. AEI did not follow a modal distribution, as significant AEI scores ranged from 4.0 – 9.1 S.D. of the gDNA error from unity, perhaps indicating a mechanism whereby a trans-acting factor with high interindividual variability interacts with the cis-acting genetic variant to produce a graded AEI.

Figure 23. Significant AEI for SF4, measured at marker SNP rs2010506. Note: Dotted lines indicate ±3 S.D. of average within-sample error.

Significant AEI was also observed for A2BP1, DLG4, EPB41L1, GAD1, GRIA2, and NRCAM, (Appendix B: Table 13) but was present in two or less samples, suggesting the occurrence of rare or uncommon variants driving allelic
expression of these genes. For A2BP1, one autistic sample displayed a 1.36-fold difference in allelic expression, which was approximately 4.3 S.D. of the error from unity for this assay. A 1.43-fold difference in allelic expression (8.6 S.D.) for one autistic sample was found for DLG4. Two autistic samples displayed low but significant AEI for EPB41L1, equaling 1.11 and 1.14-fold allelic differences (3.7 and 4.6 S.D., respectively). Two autistic samples displayed significant AEI for GAD1, with 1.77- and 2.22-fold allelic differences (3.7 and 4.6 S.D., respectively). One autistic sample, also significant for SLC1A1 AEI, displayed a 1.36-fold difference in allelic expression (3.6 S.D.) for GRIA2. Finally, one autistic sample, also AEI-positive for EPB41L1, displayed significant AEI for NRCAM, with 1.82-fold difference in allelic expression (15.6 S.D.).

4.4 Discussion

Functional variants residing in intronic regions and largely affecting mRNA expression have not been systematically examined in autism risk genes, especially in tissues relevant to the disorder from affected individuals. Considering the high percentage of disease-causing variants now found to affect mRNA processing (Cooper, 2009), it is possible that a similarly high percentage of these variants also contribute to ASD etiology. Here, we present evidence for the presence of common functional variants significantly affecting mRNA expression for SLC1A1, CNTNAP2, HNRNPA3, NOVA1, and SF4. In addition, we found evidence for less common functional SNPs that affect expression of
ASD risk genes A2BP1, DLG4, EPB41L1, GAD1, GRIA2, and NRCAM, exclusively in autistic individuals.

A previous study found evidence for the presence of genetic variants driving the expression of SLC1A1, through eQTL analyses (Wendland et al., 2009). Our results build on this finding, suggesting the presence of two cis-eQTLs, which can be used as phenotypes for uncovering the actual functional variants responsible for this expression difference. A number of studies suggest rs301430 is the responsible functional candidate, as it is associated with obsessive-compulsive disorder (Arnold et al., 2006; Dickel et al., 2006; Stewart et al., 2007) and autism (Brune et al., 2008; Gadow et al., 2010), and has an effect on expression in vitro (Wendland et al., 2009). However, 18 of 23 samples in our cohort show no difference in allelic expression measured at this SNP, arguing against any functional role for this variant. With our results indicating the presence of two potential functional regions, it is possible that rs301430 serves as the best surrogate marker for the combination of both functional alleles and the best single marker for disease risk, although this cannot be determined until the actual functional SNPs are uncovered.

Our study is the first to suggest the presence of a cis-eQTL driving expression of CNTNAP2. This gene is a member of the neurexin superfamily of cellular adhesion molecules, mediating adhesion of myelinating glia to neurons (Poliak et al., 2001). CNTNAP2 is directly implicated as a contributor to phenotypic traits in familial autism (Alarcon et al., 2008; Arking et al., 2008). A
recent study examining cortical gene expression from an evolutionary standpoint identified CNTNAP2 as differentially expressed and alternatively spliced in human versus mouse prefrontal cortex (Johnson et al., 2009), suggesting evolutionary pressures are selecting for different functions of CNTNAP2 in mice versus human. This study also found that several genes encoding CAMs that are expressed in cortical brain regions crucial to language development are subject to rapid evolutionary pressure specific to humans (Johnson et al., 2009).

A gene associated with evolution of language in humans, transcription factor forkhead box protein P2 (FOXP2) (Enard et al., 2002), directly regulates CNTNAP2 expression (Vernes et al., 2008), making any functional variants within this gene potentially relevant to speech and language development in ASD and the general population. The size of this gene (2.1 Mb) makes attempts to pinpoint loci significantly affecting expression an arduous task, but one that is necessary for future association studies where this gene is included.

Significant and frequent AEI observed for the three splicing factors, HNRNPA3, NOVA1, and SF4, requires confirmation at additional SNP locations, as each of the assays in the current study are burdened by caveats. First, these genes are members of larger families of splicing factors, with high homology in some cases. This was most evident for the biased allelic gDNA ratios observed for HNRNPA3, which if left uncorrected, would have indicated significant AEI for all samples. In this case, it is most likely that the primers for this assay were amplifying alleles for both HNRNPA3 and HNRNPA3P1 in gDNA, but only
*HNRNPA3* in cDNA, since the pseudogene is likely non-functional. Although an effort was made to correct allelic gDNA ratios for this scenario, it still does not account for the possibility of active *HNRNPA3P1* transcription. For *NOVA1*, the amplified transcript is an alternative spliceoform, and therefore not likely to be representative of expression as a whole. For *SF4*, assay quality was suspect, as SNaPshot peaks were lower than expected, indicating either inefficient PCR amplification or SNaPshot priming. Lastly, these genes are not explicitly implicated in ASD, but rather they reside in regions of the genome indicated to harbor risk in association studies (Gallagher et al., 2003; International Molecular Genetic Study of Autism Consortium, 2001; Liu et al., 2001; McCauley et al., 2005).

The remaining genes, in which AEI was significant, include the splice factor *A2BP1*, cellular adhesion-related *DLG4*, *EPB41L1*, and *NRCAM*, and the glutamate AMPA receptor 2 (*GRIA2*), each of which are compelling candidates for autism etiology. *A2BP1* was found to be partially translocated and deleted in an autistic female patient (Martin et al., 2007; Sebat et al., 2007). Apart from autism, *de novo* translocation or deletion of this gene have been associated with epilepsy and mental retardation (Bhalla et al., 2004). Many splice modulators surveyed here function in a gene dose-dependent manner (Ramocki and Zoghbi, 2008). Therefore, any mutation or epigenetic alteration that significantly changes the overall level of splicing factor protein could have pleiotropic effects on the phenotype through their interactions with other risk genes. For example,
candidate genes alternatively spliced by A2BP1 include the autism-associated
genes neuroligins 3 and 4X (NLGN3, NLGN4X), NRCAM investigate here,
neurexin 1 (NRXN1), and protocadherin 9 (PCDH9) (Zhang et al., 2008). DLG4,
perhaps better known as post-synaptic density protein 95 (PSD95), acts as an
interface between cellular adhesion molecules and glutamate receptors to
mediate excitatory versus inhibitory signaling (Prange et al., 2004). Similarly,
EPB41L1 acts as a scaffold protein that regulates activity-dependent surface
expression of AMPA and glutamate receptors (Lin et al., 2009; Shen et al., 2000)
and also stabilizes neuronal dopamine receptors (Binda et al., 2002). NRCAM
regulates neuron-to-neuron adhesion (Lane et al., 1996) and has previously been
linked to autism in numerous genetic association studies (Bonora et al., 2005;
Marui et al. 2009; Sakurai et al., 2006). GAD1 protein and mRNA expression is
dysregulated in a cell-specific manner in cerebellar cortex in autistic patients
(Blatt et al., 2001; Fatemi et al., 2002; Yip et al., 2007; Yip et al., 2008) Finally,
GRIA2 encodes for an AMPA subunit integral to synaptic plasticity (Chung et al.,
2003) and dendritic spine morphology (Passafaro et al., 2003).

Remarkably, AEI was only observed for autistic individuals for A2BP1,
DLG4, EPB41L1, GAD1, GRIA2, and NRCAM, suggesting that these individuals
harbor rare variants that significantly affect mRNA expression. As previously
discussed, it is hypothesized that autism results from highly penetrant rare
variants, rather than those more common but less penetrant (Zhao et al., 2007).
This is likely to be true for variants that change the protein coding sequence of
the risk gene, since the presence of such a variant would have an obvious effect on every single protein molecule translated from that allele. Also likely, but less obvious from sequencing studies that have examined the coding regions of risk genes, is the presence of regulatory variants having a significant effect on risk gene mRNA expression that influence ASD phenotype. For each of the genes above, it is not known how a significant reduction or increase in mRNA expression would affect autistic-like phenotypes in humans, but high conservation of the protein coding regions of these genes in vertebrates suggests the propensity to produce a severe phenotype if mutated. Proper determination of potential risk requires identification of the responsible variants and subsequently examination of their frequencies in large cohorts of autistic patients and controls.

The lack of allelic expression for many putative risk genes in this study allow us to begin eliminating genetic variants affecting mRNA expression as a mechanism for conferring risk in these genes. This includes *SLC25A12*, expression of which was previously found to be reduced in autism brain tissue (Lepagnol-Bestel et al., 2008). The marker SNP at which AEI was measured for *SLC25A12* in our study (rs11757) and those previously found to be associated with risk for ASD (rs2056202 and rs2292813) are in complete linkage disequilibrium (*D*=1) according to HapMap, as measured by SNAP (Johnson et al., 2008). Therefore, it is unlikely that rs2056202 or rs2292813 represent the actual functional SNPs associated with risk if acting at the level of RNA.
expression. It is still possible that SLC25A12, or any gene in this study, harbors uncommon functional genetic variant significantly affecting mRNA expression, since we were only able to interrogate a percentage of our population for any given gene. Our results would indicate that the allele frequency of any yet unfound variant in these genes must be quite low. For instance, the 24 samples in which AEI was measured for SLC25A12, allows for detection of alleles at a frequency of approximately 4%. More complete analysis using additional marker SNPs for each gene or high-throughput whole transcriptome sequencing will allow more definitive analysis of these genes and increase the chance of discovering rare alleles with dramatic effects on mRNA expression.

In some cases, AEI analysis can fail to reveal regulatory polymorphism, for example when alternative splicing leads to spliceoforms with similar turnover rates or transcripts with distinct transcription start sites and polyadenlyation sites. In these cases, the distinct transcripts must be assayed separately. Moreover, srSNPs can affect cellular trafficking and subcellular sequestration that would require steps to physically separate cellular components (i.e. nucleus, cytoplasm, pre- or post-synapse, etc.) prior to allelic analysis. Lastly, srSNPs can affect translation by introducing RNA secondary structures more or less permissible to the translation machinery. This can be measured at the protein level, by isolating polyribosomes that are actively translating mRNA from the gene of interest.

Advances in our understanding of RNA biology and genetic predisposition have revealed highly penetrant mechanisms whereby mRNA processing
contributes to human disease. Only by studying expression in affected tissues can one begin to estimate the frequency of disease-causing variants acting at the level of mRNA. Here, we are beginning to see compelling evidence for functional variants that affect mRNA expression that also have the potential to contribute to autistic behavioral phenotypes, as speculated for SLC1A1 and repetitive behaviors, or maybe even contribute to etiology in the cases where dramatic AEI is observed only in autistic patients.
Chapter 5: Concluding Remarks

Here, I have provided evidence for the presence of rSNPs and srSNPs in genes crucial for the manifestation of nicotine addiction, human cognition, and neuropsychiatric disorders. These SNPs have demonstrable effects on mRNA expression phenotypes, including overall expression in the case of CHRNA5 and alternative splicing in the case of HTR2A. It also appears that mRNA expression of ASD risk genes SLC1A1, CNTNAP2, HNRNPA3, NOVA1, and SF4 are driven by cis-acting functional variants, although the precise mechanisms for these genes remain to be determined. The current studies contribute to the growing body of evidence illustrating a significant effect of non-coding variants on behavioral phenotypes.

For addiction disorders, a strong environmental influence (the drug) interacts with genetic factors to drive disorder etiology. Thus, genetic risk for different drugs depends upon the systems with which the drugs interact. Not surprisingly, GWAS examining nicotine addiction finds risk associated with genetic variants in nicotinic receptor subunit genes. Biological pathways for different drugs can converge on a common pathway, as is the case with genetic risk conferred by CHRNA5 for both nicotine and cocaine addiction (Berrettini et al., 2008; Bierut et al., 2008; Grucza et al., 2008; Saccone et al., 2007; Saccone
et al., 2009a and 2009b; Sherva et al., 2008 and 2010; Stevens et al., 2008; Tobacco and Genetics Consortium, 2010; Wang et al., 2009b; Weiss et al., 2008). Some biological systems can even be ubiquitous to addiction, such as striatal dopamine pathways (Hyman et al., 2006). Similar parallels can be drawn for pharmacogenomic approaches that aim to understand drug efficacy on an interindividual basis, where genetic predisposition that predicts a clinical outcome is heavily driven by the biological targets of the drugs. In the case of drug abuse and pharmacogenomics, a known and acute environmental pressure (the drug) is imposed upon an individual to manifest an underlying genetic predisposition.

More challenging to understand from an evolutionary perspective is the manifestation of a complex genetic disorder where the responsible environmental influence is likely to be a diffuse set of factors acting over long periods of time on multiple loci. Moreover, recombination events resulting from generations of reproduction obscures our ability to pinpoint loci influenced by selection pressures over an evolutionary timescale. Thus, I speculate that it is not necessary that “functional” genetic variants in this context produce a robust phenotypic change upon which natural selection can act. Rather, as the evolutionary timescale and the size of a population increases, subtle phenotypic variation, such as RNA folding, can serve as the basis for selection.

The concept that intronic or non-protein coding variants contribute to phenotypic variability for natural selection to act upon is not novel. In fact, two recent studies reconciling the differences between robustness and evolvability
have relied upon RNA secondary structure as a model system for adaptation (Draghi et al., 2010; Wagner, 2008). In these studies, it was found that RNA secondary structure, as a phenotype, is resilient to single nucleotide changes, allowing large networks of neutral polymorphisms to accumulate. In this scenario, the neutral network allows a population of organisms access to multiple possible phenotypes needed to adapt to an environment (evolvability) while still maintaining their ability to accumulate genetic variability (robustness). This hypothesis is suitable for simple systems where the overall network of possible phenotypes and outcomes is limited; however the concept of degeneracy is necessary to drive the relationship between robustness and evolvability in complex biological systems (Whitacre, 2010). Degeneracy allows proteins that have overlapping biological functions the flexibility to alter their phenotypic nature, but with less severe consequences due to the redundancy in the system, creating a distributed robust system and greater hierarchical complexity.

Evolvability and robustness modeling provides an interesting framework for speculating about the nature of non-coding DNA in humans, and how mutations in these regions could contribute to human disease. Approximately 98% of the human genome does not code for proteins (IHGSC, 2004; Venter et al., 2001), and this relationship is recapitulated in single genes. For example, the ASD risk gene NRXN1 encompasses 1.1 million bases on chromosome 2, encoding mRNA transcripts that are approximately 7,500 bases in length, or approximately 0.7% of the total gene length. The protein coding regions of this
gene is sensitive against mutations (i.e. not robust), often producing drastic change in neuropsychiatric or cognitive phenotypes (Ching et al., 2010). Increased length between coding exons decreases the chance that any de novo mutations occur in exons, and these new mutations would serve to establish large networks of seemingly neutral polymorphisms, and high evolvability as a consequence. However, instead of the network of polymorphisms being completely neutral, they produce phenotypic changes at the RNA level that do not immediately produce changes at the behavioral level.

A comparison of human and chimpanzee first intron sequences suggest a positive correlation between evolutionary divergence and intron length (Gazave et al., 2007), perhaps suggesting a means for maintaining evolvability without compromising protein function, per se. In the face of strong recent selection pressures, as hypothesized occurred in humans to account for genetic disease risk (Crespi, 2010), selection for a modest phenotypic change at the RNA level could occur at many dynamic locations. Selection at this level would most likely not have an effect on a behavioral phenotype, due to increased redundancy as you add layers of biological complexity from genome to transcriptome to proteome, etc., until you finally reached human behavior. Only after inheriting a critical mass of these modest phenotypic changes, occurring at the RNA level and all ultimately affecting the same behavioral phenotype, does a complex genetic disorder manifest. This hypothesis provides a mechanism on which the common disease – common variant hypothesis (Hemminki et al., 2008) can
build. It also suggests that, as a phenotype used for clinical diagnosis moves away from genetics and towards biology and then behavior, the number of risk loci with the potential to have an impact increases concordantly.

Overall, the empirical data and reasoned speculation presented here suggests that uncovering phenotypes at the level of RNA expression builds a framework upon which complex human genetic disease can begin to be explained. This can be used to inform subsequent attempts to uncover genetic predisposition for any disorder and also provides new avenues for inquiry.
References


J., Krueger, R. F., Kuperman, S., Lynskey, M., Mann, K., Neuman, R. J.,
Nothen, M. M., Nurnberger, J. I., Jr., Porjesz, B., Ridinger, M., Saccone,
Rietschel, M., Goate, A. M. and Rice, J. P. (2010). A genome-
wide association study of alcohol dependence. Proc Natl Acad Sci U S A 107,
5082-5087.

receptor cell surface localization mediated by interaction with protein 4.1N.

Blasi, F., Bacchelli, E., Pesaresi, G., Carone, S., Bailey, A. J. and Maestrini, E.
and neuroligin 4 in individuals with autism from the IMGSAC collection.

Blatt, G. J., Fitzgerald, C. M., Guptill, J. T., Booker, A. B., Kemper, T. L. and
neurotransmitter receptors in autism: an autoradiographic study. J Autism
Dev Disord 31, 537-543.

Blundell, J., Tabuchi, K., Bolliger, M. F., Blaiss, C. A., Brose, N., Liu, X., Sudhof,
lacking the inhibitory synapse cell adhesion molecule neuroligin 2. Genes
Brain Behav 8, 114-126.

Bolon, I., Gouyer, V., Devouassouex, M., Vandenbunder, B., Wernert, N., Moro,
collagenase 1, and urokinase-type plasminogen activator genes in lung

Bonora, E., Lamb, J. A., Barnby, G., Sykes, N., Moberly, T., Beyer, K. S., Klauck,
S. M., Poustka, F., Bacchelli, E., Blasi, F., Maestrini, E., Battaglia, A.,
Haracopoulos, D., Pedersen, L., Isager, T., Eriksen, G., Viskum, B.,
Sorensen, E. U., Brondum-Nielsen, K., Cotterill, R., Engeland, H., Jonge,
M., Kemner, C., Steggehuis, K., Scherpenisse, M., Rutter, M., Bolton, P.
Mutation screening and association analysis of six candidate genes for
autism on chromosome 7q. Eur J Hum Genet 13, 198-207.

augmentation properties of 5-HT(2) receptor antagonists using in vivo


Davies, M. A., Conley, Y. and Roth, B. L. (2010). Functional SNPs in Genes Encoding the 5-HT2A Receptor Modify the Affinity and Potency of Several
Atypical Antipsychotic Drugs. *Biol Res Nurs*


Falvella, F. S., Galvan, A., Frullanti, E., Spinola, M., Calabro, E., Carbone, A.,


chromosomes 2q, 7q, and 16p. *Am J Hum Genet* 69, 570-581.


Li, M. D. (2006). The genetics of nicotine dependence. Curr Psychiatry Rep 8,


AMPA receptor GluR1 subunit surface expression by a 4. 1N-linked actin cytoskeletal association. *J Neurosci* 20, 7932-7940.


mutations and consequences. *Am J Hum Genet* 64, 1617-1631.


Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J. T., Abrahams, B. S., Salyakina, D., Imielinski, M., Bradfield, J. P., Sleiman, P. M., Kim, C. E.,


Appendix A: Abbreviations

5-HT – serotonin
5-HT2A – serotonin 2a receptor protein
A2BP1 – ataxin 2 binding protein 1 gene
ABI – Applied Biosystems Incorporated
AchE – acetylcholinesterase
ACTB – beta actin gene
AEI – allelic expression imbalance
AMPA – alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ANOVA – analysis of variance
ASD – autism spectrum disorder
CAM – cellular adhesion molecule
cDNA – complimentary deoxyribonucleic acid
CEPH – Centre d'Etude du Polymorphisme Humain
CEU – HapMap population of Utah residents with ancestry from northern and western Europe
CHRNA5 – cholinergic receptor, nicotinic subtype, alpha 5 subunit gene
CHRNB4 – cholinergic receptor, nicotinic subtype, beta 4 subunit gene
CNTN4 – contactin 4 gene
CNTNAP2 – contactin associated protein-like 2 gene
COGEND – Collaborative Genetic Study of Nicotine Dependence
COPD – chronic obstructive pulmonary disease
cSNP – coding single nucleotide polymorphism
C_T – cycle threshold
ddNTP - dideoxynucleotriphosphate
DLG4 – discs, large homolog 4 gene
DNA – deoxyribonucleic acid
DOI – 4-iodo-2,5-dimethoxyphenylisopropylamine
DRD2 – dopamine D2 receptor protein
EPB41L1 – erythrocyte membrane protein band 4.1-like 1 gene
eQTL – expression quantitative trait locus
Ets-1 – v-ets erythroblastosis virus E26 oncogene homolog 1 protein
FOXP2 – transcription factor forkhead box protein P2 gene
GABA – gamma-aminobutyric acid
GABRB3 – gamma-aminobutyric acid A receptor, beta 3 gene
GAD1 – glutamate decarboxylase 1 gene
gDNA – genomic deoxyribonucleic acid
GRIA2 – glutamate receptor, ionotropic, AMPA 2 gene
GRM5 – glutamate receptor, metabotropic 5 gene
GWAS – genome-wide association study
hnRNA – heterogeneous nuclear ribonucleic acid
HNRNPA3 – heterogeneous nuclear ribonucleoprotein A3 gene
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$HNRNPA3P1$</td>
<td>heterogeneous nuclear ribonucleoprotein A3 pseudogene 1</td>
</tr>
<tr>
<td>$HTR2A$</td>
<td>serotonin receptor, subtype 2A gene</td>
</tr>
<tr>
<td>IHGSC</td>
<td>International Human Genome Sequencing Consortium</td>
</tr>
<tr>
<td>LD</td>
<td>linkage disequilibrium</td>
</tr>
<tr>
<td>LSD</td>
<td>lysergic acid diethylamide</td>
</tr>
<tr>
<td>mGluR2</td>
<td>metabotropic glutamate receptor 2 protein</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NCBP2</td>
<td>nuclear cap binding protein subunit 2 gene</td>
</tr>
<tr>
<td>NEB</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>NLGN3</td>
<td>neuroligin 3 gene</td>
</tr>
<tr>
<td>NLGN4X</td>
<td>neuroligin 4, X-linked gene</td>
</tr>
<tr>
<td>NOVA1</td>
<td>neuro-oncological ventral antigen 1 gene</td>
</tr>
<tr>
<td>NRCAM</td>
<td>neuronal cellular adhesion molecule gene</td>
</tr>
<tr>
<td>NRXN1</td>
<td>neurexin 1 gene</td>
</tr>
<tr>
<td>PCDH9</td>
<td>protocadherin 9 gene</td>
</tr>
<tr>
<td>PEF</td>
<td>primer extension primer, forward gene orientation 5' → 3'</td>
</tr>
<tr>
<td>PER</td>
<td>primer extension primer, reverse gene orientation 3' → 5'</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PSD95</td>
<td>post-synaptic density 95 protein</td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative real-time polymerase chain reaction</td>
</tr>
<tr>
<td>RFLP</td>
<td>restriction fragment length polymorphism</td>
</tr>
</tbody>
</table>
RNA – ribonucleic acid
rSNP – regulatory single nucleotide polymorphism
SF4 – splicing factor 4 gene
SLC1A1 – neuronal/epithelial high affinity glutamate transporter gene
SLC1A2 – glial high affinity glutamate transporter gene
SLC25A12 – mitochondrial aspartate/glutamate exchanger gene
SLC6A1 – neurotransmitter transporter, GABA gene
SNP – single nucleotide polymorphism
SRSF – serine/arginine-rich splicing factors
srSNP – structural RNA single nucleotide polymorphism
SSRI – selective serotonin reuptake inhibitor
TSS – translation start site
UCSC – University of California, Santa Cruz
YRI – HapMap population from Yoruba in Ibadan, Nigeria
Appendix B: Tables
<table>
<thead>
<tr>
<th>Assay Type</th>
<th>SNP/Gene</th>
<th>Primer Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNaPshot*</td>
<td>rs16969968 / CHRNA5</td>
<td>F: CGATGGCCCCACTACGTGAATATTTCTTCACACGCTTCCC&lt;br&gt;R: CACCTCGACCCCCAAAAGATGTGTTCTTGTAATGTAGCGAATAG&lt;br&gt;Primer Extension: CATTGGAAGCTGCGCTC</td>
</tr>
<tr>
<td>SNaPshot*</td>
<td>rs615470 / CHRNA5</td>
<td>F: CGATGGCCCCACTACGTGATTTCCTGTACATTTTGGAAGGTATGTA&lt;br&gt;R: CACCTCGACCCCCAAAACGTGCACCGCCACATTA&lt;br&gt;Primer Extension: TATTTAAGATTAGCTGTTACTCAGTGAGT</td>
</tr>
<tr>
<td>qPCR</td>
<td>CHRNA5</td>
<td>F: ATATTTCTTCACACGCTTCCC&lt;br&gt;R: GATGTGTTCTTGTAATGTAGCGAATAG</td>
</tr>
<tr>
<td>qPCR</td>
<td>CHRNA4</td>
<td>F: GCAGGAGGCATTAGAAGGTGT&lt;br&gt;R: CACCACCATAGCCACGTACT</td>
</tr>
<tr>
<td>qPCR</td>
<td>ACTB</td>
<td>F: CCTGGACCCAGCACAAT&lt;br&gt;R: GCCGATCCACACGGAGTACT</td>
</tr>
</tbody>
</table>

* Italicized bases in primer sequence denotes target sequence for pre-amplification primers

Table 7. CHRNA5 primer sequences for quantitative AEI and qPCR assays
<table>
<thead>
<tr>
<th>Chromosome Position$^1$</th>
<th>39 HapMap SNPs in Region (minor allele frequency in CEU Samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7884131 - 78845361</td>
<td>rs880395 (0.367), rs905740 (0.367), rs7164030 (0.367)</td>
</tr>
<tr>
<td>7885228 - 78856482</td>
<td>rs3841324, rs684513 (0.217)</td>
</tr>
<tr>
<td>7885782 - 78859073</td>
<td>rs7165657 (0.000), rs7166003 (0.000)</td>
</tr>
<tr>
<td>78861175 - 78861676</td>
<td>rs7178897 (0.000), rs667282 (0.208)</td>
</tr>
<tr>
<td>78862083 - 78863528</td>
<td>rs479385 (0.000), rs16969948 (0.000), rs17486195 (0.341), rs588765 (0.339), rs6495306 (0.375)</td>
</tr>
<tr>
<td>78864617 - 78865964</td>
<td>rs16969949 (0.000), rs12903839 (0.034), rs17486278 (0.409), rs1875869 (0.009)</td>
</tr>
<tr>
<td>7886874 - 78868280</td>
<td>rs621849 (0.375), rs569207 (0.216), rs637137 (0.202), rs7180002 (0.417)</td>
</tr>
<tr>
<td>7887240 - 78874145</td>
<td>rs481134 (0.373), rs951266 (0.424), rs10519205 (0.000)</td>
</tr>
<tr>
<td>78877363 - 78878900</td>
<td>rs647041 (0.373), rs12898919 (0.133), rs12903575 (0.042), rs17408276 (0.325)</td>
</tr>
<tr>
<td>78880355 - 78881722</td>
<td>rs16969968 (0.425), rs518425 (0.250), rs514743 (0.331)</td>
</tr>
<tr>
<td>78882772 - 78884589</td>
<td>rs12899226 (0.042), rs660652 (0.325), rs472054 (0.328), rs8029939 (0.000), rs578776 (0.242)</td>
</tr>
</tbody>
</table>

$^1$ Position according to Genome Reference Consortium build 37.1 (GRCh37)

Table 8. Loci sequenced in CHRNA5 region
<table>
<thead>
<tr>
<th>SNP rs#</th>
<th>Chromosome Position¹</th>
<th>Samples Genotyped (n)</th>
<th>Minor Allele (Freq)²</th>
<th>Regression for mRNA expression (p)</th>
<th>% Samples Meeting AEI³</th>
<th>AEI Genotype Test (p)</th>
<th>LD with rs905740 D' (r²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1979907⁴</td>
<td>78842239</td>
<td>86</td>
<td>A (0.34)</td>
<td>1.13E-04</td>
<td>100</td>
<td>5.37E-12</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>rs880395⁴</td>
<td>78844356</td>
<td>87</td>
<td>A (0.33)</td>
<td>2.50E-05</td>
<td>100</td>
<td>5.37E-12</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>rs905740⁴</td>
<td>78844386</td>
<td>88</td>
<td>T (0.33)</td>
<td>2.42E-05</td>
<td>100</td>
<td>5.37E-12</td>
<td>-</td>
</tr>
<tr>
<td>rs7164030⁴</td>
<td>78844661</td>
<td>88</td>
<td>G (0.33)</td>
<td>2.42E-05</td>
<td>100</td>
<td>5.37E-12</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>rs2036527</td>
<td>78851615</td>
<td>90</td>
<td>A (0.35)</td>
<td>5.75E-04</td>
<td>55</td>
<td>6.09E-01</td>
<td>1.00 (0.27)</td>
</tr>
<tr>
<td>rs3841324⁵</td>
<td>78857813:78857834</td>
<td>89</td>
<td>S (0.35)</td>
<td>1.81E-03</td>
<td>86</td>
<td>4.48E-03</td>
<td>0.83 (0.64)</td>
</tr>
<tr>
<td>rs667282</td>
<td>78863472</td>
<td>87</td>
<td>C (0.29)</td>
<td>8.95E-01</td>
<td>30</td>
<td>7.13E-02</td>
<td>0.54 (0.06)</td>
</tr>
<tr>
<td>rs17486195</td>
<td>78865197</td>
<td>88</td>
<td>G (0.30)</td>
<td>5.92E-03</td>
<td>55</td>
<td>8.29E-01</td>
<td>0.89 (0.17)</td>
</tr>
<tr>
<td>rs680244</td>
<td>78871288</td>
<td>91</td>
<td>T (0.38)</td>
<td>2.71E-05</td>
<td>79</td>
<td>2.73E-03</td>
<td>0.81 (0.52)</td>
</tr>
<tr>
<td>rs621849</td>
<td>78872861</td>
<td>90</td>
<td>G (0.37)</td>
<td>3.45E-06</td>
<td>83</td>
<td>7.16E-04</td>
<td>0.84 (0.57)</td>
</tr>
<tr>
<td>rs692780</td>
<td>78876505</td>
<td>91</td>
<td>G (0.30)</td>
<td>6.75E-05</td>
<td>77</td>
<td>6.77E-03</td>
<td>0.83 (0.61)</td>
</tr>
<tr>
<td>rs17408276</td>
<td>78881618</td>
<td>88</td>
<td>C (0.27)</td>
<td>1.38E-02</td>
<td>77</td>
<td>2.69E-02</td>
<td>0.83 (0.53)</td>
</tr>
<tr>
<td>rs16969968⁶</td>
<td>78882925</td>
<td>87</td>
<td>A (0.30)</td>
<td>1.46E-02</td>
<td>58</td>
<td>5.83E-01</td>
<td>0.89 (0.17)</td>
</tr>
<tr>
<td>rs514743</td>
<td>78884227</td>
<td>91</td>
<td>A (0.34)</td>
<td>1.21E-03</td>
<td>79</td>
<td>1.11E-02</td>
<td>0.71 (0.49)</td>
</tr>
<tr>
<td>rs615470</td>
<td>78885988</td>
<td>89</td>
<td>T (0.37)</td>
<td>4.76E-03</td>
<td>76</td>
<td>5.65E-02</td>
<td>0.71 (0.42)</td>
</tr>
<tr>
<td>rs578776</td>
<td>78888400</td>
<td>91</td>
<td>A (0.35)</td>
<td>9.40E-01</td>
<td>29</td>
<td>3.55E-03</td>
<td>0.66 (0.11)</td>
</tr>
</tbody>
</table>

¹ Position according to Genome Reference Consortium build 37.1 (GRCh37)
² Allele frequency calculated in our PFC sample population
³ To meet criterion sample must be heterozygous in AEI positive samples, homozygous in AEI negative samples
⁴ proposed enhancer SNPs
⁵ 22-bp deletion variant
⁶ nonsynonymous SNP

Table 9. SNP associations with overall CHRNA5 expression and AEI, and LD with rs905740
### Table 10. Linear regression for *CHRNA5* mRNA expression across genotype with race considered as a covariate and separate regression analyses within each race group.

<table>
<thead>
<tr>
<th>Polymorphism rs#</th>
<th>Bonferroni-corrected p-values</th>
<th>Uncorrected p-values within each race cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall regression with race as covariate (n=92)¹</td>
<td>European-American (n=59)¹</td>
</tr>
<tr>
<td>rs1979907</td>
<td>1.15E-03</td>
<td>0.001228</td>
</tr>
<tr>
<td>rs880395</td>
<td>2.52E-04</td>
<td>0.000249</td>
</tr>
<tr>
<td>rs905740</td>
<td>2.19E-04</td>
<td>0.000236</td>
</tr>
<tr>
<td>rs7164030</td>
<td>2.19E-04</td>
<td>0.000236</td>
</tr>
<tr>
<td>rs2036527</td>
<td>1.71E-02</td>
<td>0.036361</td>
</tr>
<tr>
<td>rs3841324</td>
<td>1.17E-02</td>
<td>0.0000186</td>
</tr>
<tr>
<td>rs667282</td>
<td>1.00E+00</td>
<td>0.641643</td>
</tr>
<tr>
<td>rs17486195</td>
<td>1.39E-01</td>
<td>0.023222</td>
</tr>
<tr>
<td>rs680244</td>
<td>1.68E-04</td>
<td>0.0000869</td>
</tr>
<tr>
<td>rs621849</td>
<td>2.57E-05</td>
<td>0.00000724</td>
</tr>
<tr>
<td>rs692780</td>
<td>2.71E-04</td>
<td>0.0000211</td>
</tr>
<tr>
<td>rs17408276</td>
<td>6.65E-02</td>
<td>0.002868</td>
</tr>
<tr>
<td>rs16969968</td>
<td>3.96E-01</td>
<td>0.020725</td>
</tr>
<tr>
<td>rs514743</td>
<td>7.89E-03</td>
<td>0.0000447</td>
</tr>
<tr>
<td>rs615470</td>
<td>7.26E-02</td>
<td>0.000791</td>
</tr>
<tr>
<td>rs578776</td>
<td>1.00E+00</td>
<td>0.967985</td>
</tr>
</tbody>
</table>

¹Number of subjects genotyped. Genotype assignment was not successful in all subjects, depending on the SNP analyzed; this difference accounts for small changes in p-values between SNPs in complete LD.
<table>
<thead>
<tr>
<th>Assay Type*</th>
<th>SNP/Gene</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNaPshot</td>
<td>rs6313/HTR2A</td>
<td>F: CTCAACTACGAACCTCCCTAATGCAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: TTGGTTCGATTTTCAGAGTCGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PER: CATCAGAAGTGTAGCTTCTCC</td>
</tr>
<tr>
<td>SNaPshot</td>
<td>rs6314/HTR2A</td>
<td>F: GCAAGATGCCAAGACACAGATAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: TCACACACAGCCTACCTTTTCAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PEF: TGGTTGCTCTAGGAAAAGCAG</td>
</tr>
<tr>
<td>SNaPshot</td>
<td>rs3803189/HTR2A</td>
<td>F: GATGACATGGGATTGAGTTGTTAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GCAATACAGATTTTATAACACTGAACCTTAGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PEF: CCATTATATTCAATAAAATTTTCACTATT</td>
</tr>
<tr>
<td>SNaPshot</td>
<td>rs6312/HTR2A</td>
<td>F: CTGTGAGAGATGCAGCGAGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: AAGCATGATTTCACAAACGGAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PEF: GAATAACAAATGTATCTCATGTG</td>
</tr>
<tr>
<td>qPCR</td>
<td>Total</td>
<td>F: GCAAGATGCCAAGACACAGATAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: TCACACACAGCCTACCTTTTCAT</td>
</tr>
<tr>
<td>qPCR</td>
<td>E2⁺</td>
<td>F: CTGTGAGAGATGCAGCGAGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: TCGGGAAGATAAATGTACATGTG</td>
</tr>
<tr>
<td>qPCR</td>
<td>E2⁻</td>
<td>F: CTGTGAGAGATGCAGCGAGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: GCCACCGGTACCATTGTC</td>
</tr>
<tr>
<td>qPCR</td>
<td>E2lr</td>
<td>F: CTGTGAGAGATGCAGCGAGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: CAGACCAGTTTTTTCATTGTC</td>
</tr>
</tbody>
</table>

* Reverse primers from each SNaPshot assay were used for gene-specific priming during cDNA synthesis

Table 11. *HTR2A* primers used for SNaPshot, qPCR, and cDNA synthesis
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>SNP</th>
<th>Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CNTNAP2</strong></td>
<td>contactin associated protein-like 2</td>
<td>rs2530312</td>
<td>F: TTCCATCCAACTACACTGCTGT&lt;br&gt;R: CTACACTTCTCCATCCTCTGGGA&lt;br&gt;PER: GATTCCAGGGATTTTTCTTCA</td>
</tr>
<tr>
<td><strong>CNTN4</strong></td>
<td>contactin 4</td>
<td>rs10888</td>
<td>F: TTGCTATAGTTTGTCAATTTCGTTACA&lt;br&gt;R: CCGGTTACCACCCTAGTTTTTACA&lt;br&gt;PER: CAAATGGAATCCCGATTATG</td>
</tr>
<tr>
<td><strong>DLG4</strong></td>
<td>discs, large homolog 4</td>
<td>rs17203281</td>
<td>F: CTGCCATGCCCCTGAAGAAT&lt;br&gt;R: CCGAAGGTCTGGATCTTG&lt;br&gt;PER: CAAATGGAATCCCGATTATG</td>
</tr>
<tr>
<td><strong>EPB41L1</strong></td>
<td>erythrocyte membrane protein band 4.1-like 1</td>
<td>rs3813918</td>
<td>F: GAGCACAAGACACACCTCAGA&lt;br&gt;R: GGCCCATAAATCTCACAAGCT&lt;br&gt;PER: GGATTTTGCACCTGGCTT</td>
</tr>
<tr>
<td><strong>NRCAM</strong></td>
<td>neuronal cell adhesion molecule</td>
<td>rs2072546</td>
<td>F: TGACATCGATAAAGACCCCTCTTGG&lt;br&gt;R: TGTATAGAAACTGCGACTCAG&lt;br&gt;PER: GAGTCTATCAGTGTACAGCAAGGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs1269621</td>
<td>F: AGAAAAACCCTTGCAAGTCTATCTATG&lt;br&gt;R: CTTTAACCTAAAGAAATGGATATGTGGG&lt;br&gt;PER: TGGATGGCTCCTAATGC</td>
</tr>
</tbody>
</table>

Table 12. Primers for ASD genes investigated in BA10 cortex samples using SNaPshot.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>SNP</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>PEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCDH9</td>
<td>protocadherin 9</td>
<td>rs9571740</td>
<td>F: CTCCAGAGTATTTGACAGAGAGAACA</td>
<td>R: CAGAACAGTAACAAATCACAGCCG</td>
<td>TTACAGTAACTGCCAGGGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glutamate/GABA Signaling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABRB3</td>
<td>gamma-aminobutyric acid (GABA) A receptor, beta 3</td>
<td>rs2017247</td>
<td>F: CTGGCTGTCACTGAGCACCA</td>
<td>R: GCAGTAGAAGAGGTAAATCAATCAGGGA</td>
<td>ATTTAATTCACCTGGGCAG</td>
</tr>
<tr>
<td>GAD1</td>
<td>glutamate decarboxylase 1</td>
<td>rs3749034</td>
<td>F: AACTAGCGAGAACGAGGAAGCA</td>
<td>R: TTCCTCTGCACAGCGCC</td>
<td>ATTTAATTCACCTGGGCAG</td>
</tr>
<tr>
<td>GRIA2</td>
<td>glutamate receptor, ionotropic, AMPA 2</td>
<td>rs6855973</td>
<td>F: GTGAAAAAGAAATGACTATGAC</td>
<td>R: GCCGACAGAAGAAAATTGCAGATAAGCTA</td>
<td>ATTTAATTCACCTGGGCAG</td>
</tr>
<tr>
<td>GRM5</td>
<td>glutamate receptor, metabotropic 5</td>
<td>rs3462</td>
<td>F: GGGCTTAGAAACATGGGGCTT</td>
<td>R: ACACGCACTGCTTCCCTAGA</td>
<td>ATTTAATTCACCTGGGCAG</td>
</tr>
<tr>
<td>SLC1A1</td>
<td>neuronal/epithelial high affinity glutamate transporter</td>
<td>rs301430</td>
<td>F: CCAGGTGGACAAGAGAGGCAC</td>
<td>R: TGCAATAACACCGCTGCA</td>
<td>ATTTAATTCACCTGGGCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs2228622</td>
<td>F: GAGCATCAAGGCCCTGTCAGTCA</td>
<td>ATTTAATTCACCTGGGCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rs3087879</td>
<td>R: AAGATTAGTTAGAACAAGTCCTCTCTG</td>
<td>ATTTAATTCACCTGGGCAG</td>
</tr>
</tbody>
</table>

Continued
Table 12. Primers for ASD genes investigated in BA10 cortex samples using SNaPshot

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>rs</th>
<th>Forward Primer (F)</th>
<th>Reverse Primer (R)</th>
<th>Permutation (PER)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC1A2</td>
<td>glial high affinity glutamate transporter</td>
<td>rs752949</td>
<td>F: CGAAGAAAGTCCTGGTTGCAC</td>
<td>R: TCCGGCACCTCAGTGCAAGCA</td>
<td>PER: CGTTGGCCTCCTCCTGTC</td>
</tr>
<tr>
<td>SLC25A12</td>
<td>mitochondrial aspartate/glutamate exchanger</td>
<td>rs11757</td>
<td>F: GTGTAGACCTCCTCTGTTATCA</td>
<td>R: CTTGGCCTTGGCTTACTGTG</td>
<td>PEF: GCTCGTGCTTGCA</td>
</tr>
<tr>
<td>SLC6A1</td>
<td>neurotransmitter transporter, GABA</td>
<td>rs1062246</td>
<td>F: CCAAGATTCTTGGCTTACATATCAAAA</td>
<td>R: GCCTCAAGTTTCTACGATATACATAAAAA</td>
<td>PEF: ACATGCACATCCATTCCTCA</td>
</tr>
<tr>
<td>A2BP1</td>
<td>ataxin 2 binding protein 1</td>
<td>rs881562</td>
<td>F: CCGAGGCCTGAGTTGCA</td>
<td>R: CCAAACTACGGAACCTCAGAAA</td>
<td>PER: ATATCTGAGGTATAAGATAAATGT</td>
</tr>
<tr>
<td>HNRNPA3</td>
<td>heterogeneous nuclear ribonucleoprotein A3</td>
<td>rs8470</td>
<td>F: CCAAGCATGGAAATTGGTCTTCAC</td>
<td>R: GCCTTTTCTACGCTTTGCAC</td>
<td>PER: ATGTCTTGGAAACTGCAGTAT</td>
</tr>
<tr>
<td>NCBP2</td>
<td>nuclear cap binding protein subunit 2</td>
<td>rs11550120</td>
<td>F: ACAGGACAGCAGACTTCCCAT</td>
<td>R: GCTTTAGTATAACACTGACATCTTTTGG</td>
<td>PER: AATACTCAAGTTTCCGGCTCA</td>
</tr>
<tr>
<td>NOVA1</td>
<td>neuro-oncological ventral antigen 1</td>
<td>rs1245213</td>
<td>F: TTTGTTGGATTTGTTTCTCTGTCA</td>
<td>R: CCTTTTTCTTATCTTGGTTATATATTTCA</td>
<td>PER: GATTAAATCGATAGCAAAGCTAAA</td>
</tr>
<tr>
<td>SF4</td>
<td>splicing factor 4</td>
<td>rs2010506</td>
<td>F: GAATGGTCCGGGCTGCA</td>
<td>R: TGCTACCCGCTGCTAACAG</td>
<td>PEF: CGTTGGCAGGCTGAGGCTG</td>
</tr>
<tr>
<td>Gene symbol</td>
<td>C&lt;sub&gt;T&lt;/sub&gt; Average</td>
<td>SNPs / Samples screened</td>
<td>AEI-positive*</td>
<td>AEI Range*</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------</td>
<td>-------------------------</td>
<td>--------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td><strong>Cellular adhesion and synapse structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNTNAP2</td>
<td>25.4</td>
<td>1/25</td>
<td>4</td>
<td>0.81 – 1.22</td>
<td></td>
</tr>
<tr>
<td>CNTN4</td>
<td>33.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLG4</td>
<td>23.5</td>
<td>1/19</td>
<td>1</td>
<td>0.70 – 1.10</td>
<td></td>
</tr>
<tr>
<td>EPB41L1</td>
<td>27.5</td>
<td>1/16</td>
<td>2</td>
<td>0.93 – 1.13</td>
<td></td>
</tr>
<tr>
<td>NRCAM</td>
<td>24.6</td>
<td>3/39</td>
<td>1</td>
<td>0.77 – 1.82</td>
<td></td>
</tr>
<tr>
<td>PCDH9</td>
<td>26</td>
<td>1/29</td>
<td>0</td>
<td>0.75 – 1.22</td>
<td></td>
</tr>
<tr>
<td><strong>Glutamate and GABA signaling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABRB3</td>
<td>29.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD1</td>
<td>28.2</td>
<td>1/16</td>
<td>2</td>
<td>0.45 – 1.08</td>
<td></td>
</tr>
<tr>
<td>GRIA2</td>
<td>26.6</td>
<td>1/13</td>
<td>1</td>
<td>0.90 – 1.36</td>
<td></td>
</tr>
<tr>
<td>GRM5</td>
<td>28.7</td>
<td>1/10</td>
<td>0</td>
<td>0.89 – 1.11</td>
<td></td>
</tr>
<tr>
<td>SLC1A1</td>
<td>28.6</td>
<td>3/42</td>
<td>14</td>
<td>0.64 - 2.51</td>
<td></td>
</tr>
<tr>
<td>SLC1A2</td>
<td>29.1</td>
<td>1/17</td>
<td>0</td>
<td>0.89 – 1.06</td>
<td></td>
</tr>
<tr>
<td>SLC25A12</td>
<td>25.2</td>
<td>1/24</td>
<td>0</td>
<td>0.89 – 1.21</td>
<td></td>
</tr>
<tr>
<td>SLC6A1</td>
<td>28.6</td>
<td>1/24</td>
<td>0</td>
<td>0.91 – 1.01</td>
<td></td>
</tr>
<tr>
<td><strong>Alternative splicing and RNA processing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2BP1</td>
<td>20.7</td>
<td>1/22</td>
<td>1</td>
<td>0.81 – 1.36</td>
<td></td>
</tr>
<tr>
<td>HNRNPA3</td>
<td>22.7</td>
<td>1/20</td>
<td>4</td>
<td>0.84 – 1.54</td>
<td></td>
</tr>
<tr>
<td>NCBP2</td>
<td>23.8</td>
<td>1/20</td>
<td>0</td>
<td>0.92 – 1.08</td>
<td></td>
</tr>
<tr>
<td>NOVA1</td>
<td>28.8</td>
<td>1/16</td>
<td>5</td>
<td>0.97 – 1.45</td>
<td></td>
</tr>
<tr>
<td>SF4</td>
<td>28.9</td>
<td>1/11</td>
<td>6</td>
<td>0.46 – 1.69</td>
<td></td>
</tr>
</tbody>
</table>

* - Samples are designated AEI-positive if their AEI is outside of ± 3 standard deviations of the assay for the entire population

Table 13. Expression level (C<sub>T</sub>) and AEI for autism risk genes surveyed in BA10 cortex samples
Appendix C: Figures
Figure 24. Alignment of vertebrate 5-HT2A translation start sites (TSS), exon 2 splicing, and rs6313. The E2tr splice site corresponds to the conserved TSS for Fugu, Stickleback, Medaka, and Lamprey.