From Linkage to GWAS: A Multifaceted Exploration of the Genetic Risk for Alcohol Dependence

Amy Adkins

Virginia Commonwealth University

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

Part of the Medical Genetics Commons

© The Author

Downloaded from https://scholarscompass.vcu.edu/etd/2920

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.
From Linkage to GWAS:
A Multifaceted Exploration of the Genetic Risk for Alcohol Dependence

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Amy Elizabeth Adkins, B.S.
Department of Human and Molecular Genetics
Virginia Commonwealth University

Director: Brien P. Riley, Ph.D.
Associate Professor, Departments Psychiatry and Human and Molecular Genetics

Virginia Commonwealth University
Richmond, Virginia
December, 2012
ACKNOWLEDGMENT

I have been lucky enough to be surrounded by wonderful friends, family and colleagues that have supported me professionally and personally during my tenure in graduate school. My advisor, Brien Riley, has been exceedingly kind and generous with his time. He has taught me what it means to be a great teacher and kindled a desire in me to pursue the same. Danielle Dick, Jill Bettinger and Mike Grotewiel have been incredibly supportive committee members who helped me develop as a scientist and as an individual. I sincerely appreciate the time they spent with me and their advice. Brion Maher and Darlene Kertes have been integral members of my committee and their kindness and expertise have made this final dissertation possible. Thank you.

The number of people in the Department of Human and Molecular Genetics and the Virginia Institute for Psychiatric and Behavioral Genetics who have helped and supported me makes it impossible to be exhaustive in my acknowledgments. There are several people who stand out, however. I would like to thank Mike Neale for funding my work with his training grant. Rita Shiang has been a constant source of optimism, support and sage advice. My research was completed thanks to the patience, wisdom, and generosity of Fazil Aliev, Steve Aggen and Vlady Vladimirov. I cannot express how grateful I am for their time and all that I have learned from them. Also, many thanks to Alexis Edwards,
Brandon Wormley, Gursharan Kalsi, Elizabeth Prom-Wormley, Jackie Meyers, Jessica Salvatore, Jia Yan, Shaun Latendresse, Tim Bigdeli and Vernell Williamson for helpful advice, friendship and expert analysis.

I would especially like to thank my friends and family for their endless support and love. I was fortunate enough to meet two wonderful women while in graduate school, Dawn Thiselton and Laura Hack, who inspire me and whose friendship brings me much joy. Thanks to all the many other friends I've made at VCU, including Aaron, Anjana, Ankita, Belal, Brandon, Chris, Dave, Erik, Jim, Kimmie, Megan, Mohammed, Rachel and Roseanne. Last, but certainly not least, I would like to express my gratitude and love for my large, loud and wonderful family. In particular, my grandparents, C.E. and L.R, mother, Pam, and brother, David, are the reasons that I was able to pursue, attend and finish graduate school. Any success I achieve is due to their love and support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>vii</td>
</tr>
<tr>
<td>Abstract</td>
<td>x</td>
</tr>
<tr>
<td>Chapter One. Global Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter Two. Investigating the Role of <em>GPSM1</em> in Human Alcohol Dependence and Craving Behaviors</td>
<td>21</td>
</tr>
<tr>
<td>Chapter Three. Association of <em>GRM3</em> and <em>KIAA1324L</em> with Alcohol Dependence in the Irish Affected Sib Pair Study of Alcohol Dependence</td>
<td>39</td>
</tr>
<tr>
<td>Chapter Four. A Genome-wide Association Study of Initial Sensitivity in the Irish Affected Sib Pair Study of Alcohol Dependence</td>
<td>62</td>
</tr>
<tr>
<td>Chapter Five. A Genome-wide Association Study of Maximum Alcohol Consumption in the Irish Affected Sib Pair Study of Alcohol Dependence</td>
<td>80</td>
</tr>
<tr>
<td>Chapter Six. Examination of the Shared Variation Between Alcohol Dependence, Major Depression and Neuroticism</td>
<td>94</td>
</tr>
<tr>
<td>Chapter Seven. Global Conclusions and Future Directions</td>
<td>111</td>
</tr>
<tr>
<td>List of References</td>
<td>115</td>
</tr>
<tr>
<td>Clarification of Contributions</td>
<td>147</td>
</tr>
<tr>
<td>Author’s Vita</td>
<td>149</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1.1. Survey of published linkage studies with LOD scores > 2 ........................................4
1.2. Survey of published GWAS for AD-related phenotypes .........................................14
2.1. Phenotype distributions in the case-control analyses ..............................................27
2.2. Phenotype distributions in the case-only analyses ...................................................28
2.3. Association results for GPSM1 in the IASPSAD and COGA case-control samples .................................................................31
2.4. Association results for GPSM1 within IASPSAD and COGA AD cases ..............32
2.5. Gene set of GPSM1-related genes ........................................................................35
3.1. Validation of pooled GWAS results by individual genotyping in the IASPSAD .................................................................48
3.2. Case-control association results for AD in the IASPSAD and COGA ................50
3.3. Case-control association results for AD symptom count in the IASPSAD, COGA and MGS samples ..................................................................................51
3.4. Gene expression differences between chronic alcoholics and controls ..............53
3.5. Association results for allele-specific expression differences .............................53
4.1. SNPs associated with ISENS surviving multiple testing correction ..................73
4.2. Results for ISENS gene-based association ..........................................................74
5.1. SNPs associated with MAX24 surviving multiple testing correction...............88
5.2. Results for MAX24 gene-based association ..........................................................88
6.1. Correlation between phenotypes in the IASPSAD and COGA .........................102
6.2. Factor loadings on the common factor ................................................................103
6.3. Invariance modeling results ..................................................................................108
6.4. Factor loadings using the partial invariance model with AD variant ..........108
# LIST OF FIGURES

2.1. Eigenvalues and scree plots for principal components analysis ................................................. 30

2.2. Reported CNVs in, and proximal to, GPSM1 ................................................................. 38

3.1. Significant GRM3 gene expression differences in human chronic alcoholics and control brain tissues ........................................................................................................... 54

3.2. LD in the associated region of chromosome 7 in IASPSAD controls ............................... 56

3.3. LD in the associated region of chromosome 7 in COGA controls of European descent ............................................................................................................................... 56

3.4. LD in the associated region of chromosome 7 in MGS2 controls of European descent ............................................................................................................................... 57

3.5. ENCODE graphic for the associated region on chromosome 7 .............................................. 61

4.1. Distribution of ISENS values in the IASPSAD ................................................................. 71

4.2. QQ plot for ISENS ........................................................................................................... 72

4.3. Manhattan plot for ISENS ............................................................................................. 72

5.1. Distribution of MAX24 values in the IASPSAD ................................................................. 86

5.2. QQ plot for MAX24 ......................................................................................................... 87

5.3. Manhattan plots for MAX24 .......................................................................................... 87

6.1. Phenotype distributions in the IASPSAD and COGA ......................................................... 104
LIST OF ABBREVIATIONS

AD .....................alcohol dependence
ADH .....................alcohol dehydrogenase
ALDH ..................aldehyde dehydrogenase
ALIGATOR .............Association List Go AnnoTatOR
ANOVA ...............analysis of variance
ARP .....................alcohol-related phenotype
BLA .....................basolateral nucleus of the amygdala
BMP ....................bone morphogenetic protein
cAMP ...................cyclic adenosine monophosphate
CNA ....................central nucleus of the amygdala
CDCV ..................common disease - common variant
CDRV ..................common disease - rare variant
CFI .....................Comparative Fit Index
CIDI-SF .............Composite International Diagnostic Interview-Short Form
CIDR ..................Center for Inherited Disease Research
CNS .....................central nervous system
CNV .....................copy number variation
COGA ..................Collaborative Study on the Genetics of Alcoholism
COGEND ..........Collaborative Genetic Study of Nicotine Dependence
CTX ...................superior frontal cortex
dbGaP .................database of genotypes and phenotypes
NCS .................National Comorbidity Survey
NEO-FFI ............NEO Five-Factor Inventory
NESARC ............National Epidem. Survey on Alcohol & Related Conditions
NIMH ..............National Institute of Mental Health
NLAES .............National Longitudinal Alcohol Epidemiological Survey
NSW TRC ..........New South Wales Tissue Resource Center
OZALC .............The Australian twin-family study of alcohol use disorder
PCA ...............principal components analysis
PMI ..................postmortem interval
qPCR...............quantitative polymerase chain reaction
QTL .................quantitative trait loci
RDCA ..............Research Diagnostic Criteria for Alcoholism
RIN ..................RNA integrity number
RMSEA ............Root Mean Square Error Approximation
SAGE ..............Study of Addiction: Genetics and Environment
SNP ...............single nucleotide polymorphism
SRE ...............Self-Rating for the Effects of Alcohol
SSAGA ............Semi-Structured Assessment of the Genetics of Alcoholism
TGEN ..............Translational Genomics Research Institute
TLI ..................Tucker-Lewis Index
tSNP ..............tagging single nucleotide polymorphism
VEGAS ...........Versatile Gene-Based test for GWAS Studies
WLSMV ............weighted least squares means and variances adjusted
ABSTRACT

FROM LINKAGE TO GWAS: A MULTIFACETED EXPLORATION OF THE GENETIC RISK FOR ALCOHOL DEPENDENCE

By Amy Elizabeth Adkins, B.S.

A dissertation submitted in partial fulfillment of the requirements for the degree Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2012.

Major Director: Brien P. Riley, Ph.D.

Associate Professor, Departments of Psychiatry and Human and Molecular Genetics

Family, twin and adoption studies consistently suggest that genetic factors strongly influence the risk for alcohol dependence (AD). Although the literature supports the role of genetics in AD, identification of specific genes contributing to the etiology of AD has proven difficult. These difficulties are due in part to the complex set of risk factors contributing to the development of AD. These risk factors include comorbidities with other clinical diagnoses and behavioral
phenotypes (e.g., major depression), physiological differences that contribute to the differences between people in their level of response to ethanol (e.g., initial sensitivity) and finally the large number of biological pathways targeted by and involved in the processing of ethanol. These complexities have probably contributed to the limited success of linkage and candidate gene association studies in finding genes underlying AD. The powerful and unbiased genome-wide association study (GWAS) offers promise in the study of complex diseases. However, due to the complexities of known risk factors, GWAS data has yet to provide consistent, replicable results. In light of these difficulties, this dissertation has five specific aims which attempt to investigate genetic risk loci for AD and related phenotypes through improved methods for candidate gene selection, analysis of a pooled genome-wide association study, genome-wide analyses of initial sensitivity and maximum alcohol consumption in a twenty-four hour period and finally, creation of a multivariate AD/internalizing phenotype.
CHAPTER ONE

Global Introduction

Alcohol dependence (AD) is a debilitating and chronic disorder with lifetime prevalence estimates in the United States of 20% in males and 10% in females (Kessler, Crum et al. 1997, Grant 1997). Harmful alcohol use is a significant global health burden and is estimated to contribute to 2.5 million deaths worldwide each year (World Health Organization 2011). Family, twin and adoption studies consistently suggest that genetic factors strongly influence the risk for AD. AD is a dichotomous clinical diagnosis, defined by the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) as the presence of 3 of the following 7 symptoms over the same period of 12 months: inability to quit drinking, drinking more than intended, withdrawal, spending excessive amounts of time related to drinking, tolerance, negative consequences of drinking on social, work and recreational activities and finally, continued drinking despite negative consequences (American Psychiatric Association 2000). Although the literature supports the role of genetic variation in the etiology of AD, identification of specific genes contributing to the development and maintenance of AD has proven difficult. These difficulties are due in part to the complex etiology of this pathological behavior and the likely contribution to risk for AD from numerous genes in multiple distinct pathways and networks. An analysis of the genetic epidemiology of AD and the complexity of the AD phenotype highlights the diverse and complicated nature of its genetic risk factors.
Part I: The genetic epidemiology of AD

Family, Adoption, and Twin Studies

AD runs in families. Early studies documented that alcoholics were more likely to have alcoholic family members in comparison to non-alcoholics or the population (reviewed in (Cotton 1979)). Possible hypotheses for the transmission of the disease included both heritable and environmental risk. Multiple reports looking at adopted male children of alcoholics provided evidence for genetic susceptibility for AD and alcohol abuse as these males had increased rates of AD even though removed from their alcoholic family environment (Goodwin, Schulsinger et al. 1973, Cloninger, Bohman et al. 1981, Sigvardsson, Bohman et al. 1996, Cadoret, Troughton et al. 1987). Other studies examined the rates of alcoholism in female and male adoptees and found correlations between alcohol abuse in the adoptees and problem drinking (Cadoret, O'Gorman et al. 1985) or alcoholism (Bohman, Cloninger et al. 1987) in the biological parents.

twin studies provided the first statistical estimates of the genetic contribution to the risk of AD. A number of different approaches have been used to elucidate the genes underlying this risk.

**Linkage studies**

Linkage analysis investigates markers across the genome to locate chromosomal regions segregating within families in affected (but not unaffected) individuals with the phenotype of interest (Hirschhorn 2005). There are numerous published linkage studies for AD and related phenotypes. Table 1.1 summarizes linkage studies in the literature with published LOD scores >2. In a 2009 review, Li and Burmeister (Li, Burmeister 2009) noted that chromosomes 2, 4, 5, 7, 9, 10 and 13 had either significant (LOD > 3.6) or suggestive (LOD between 2.2 and 3.6) peaks related to alcoholism (Lander, Kruglyak 1995). Major linkage studies for AD have been performed on Mission Indian and Southwest American Indian tribes from the United States (Ehlers, Gilder et al. 2004, Long, Knowler et al. 1998), multiplex families in the Pittsburgh area (Hill, Shen et al. 2004), the University of California, San Francisco family alcoholism study (Gizer, Ehlers et al. 2011), the Collaborative Study on the Genetics of Alcoholism (COGA) (Reich, Edenberg et al. 1998, Foroud, Edenberg et al. 2000, Corbett, Saccone et al. 2005), the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) (Prescott, Sullivan et al. 2006) and a four site study in the United States ((Panhuysen, Kranzler et al. 2010); European American subset and (Gelernter, Kranzler et al. 2009); African American subset).
### Table 1.1. Survey of Published Linkage Studies with LOD scores >2.

<table>
<thead>
<tr>
<th>Study</th>
<th>Phenotype</th>
<th>Sample (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corbett, 2005</td>
<td>AD</td>
<td>COGA (1464 families)</td>
</tr>
<tr>
<td>Foroud, 2000</td>
<td>AD</td>
<td>COGA (262 families)</td>
</tr>
<tr>
<td>Gelernter, 2009</td>
<td>AD</td>
<td>US/African American (238 families)</td>
</tr>
<tr>
<td>Gizer, 2011</td>
<td>AD</td>
<td>UCSF (556 probands, 1080 1st deg relatives)</td>
</tr>
<tr>
<td>Hill, 2004</td>
<td>AD</td>
<td>Pittsburgh (330 ppl)</td>
</tr>
<tr>
<td>Long, 1998</td>
<td>AD</td>
<td>Southwestern American Indians (172 sib pairs)</td>
</tr>
<tr>
<td>Panhuysen, 2005</td>
<td>AD</td>
<td>US/Europ. American (95 families)</td>
</tr>
<tr>
<td>Prescott, 2006</td>
<td>AD</td>
<td>Irish (474 families)</td>
</tr>
<tr>
<td>Reich, 1998</td>
<td>AD</td>
<td>COGA (105 families)</td>
</tr>
<tr>
<td>Bergen, 2003</td>
<td>Max. alcohol consumption</td>
<td>Framingham Heart Study (2849 ppl)</td>
</tr>
<tr>
<td>Dick, 2002</td>
<td>Quantitative alcohol-related</td>
<td>COGA (262 families)</td>
</tr>
<tr>
<td>Ehlers, 2004</td>
<td>Drinking severity</td>
<td>Mission Indians (243 ppl)</td>
</tr>
<tr>
<td></td>
<td>Withdrawal</td>
<td>Mission Indians (243 ppl)</td>
</tr>
<tr>
<td>Ehlers, 2005</td>
<td>Alcohol craving</td>
<td>Mission Indians (885 ppl)</td>
</tr>
<tr>
<td>Ehlers, 2010</td>
<td>Initial sensitivity</td>
<td>American Indians (381 ppl)</td>
</tr>
<tr>
<td>Foroud, 1998</td>
<td>Severe drinking</td>
<td>COGA (105 families)</td>
</tr>
<tr>
<td>Gizer, 2011</td>
<td>Severe drinking symptoms</td>
<td>UCSF (556 probands, 1080 1st deg relatives)</td>
</tr>
<tr>
<td>Kuo, 2006</td>
<td>Age of AD onset</td>
<td>Irish (474 families)</td>
</tr>
<tr>
<td></td>
<td>Initial sensitivity</td>
<td>Irish (474 families)</td>
</tr>
<tr>
<td></td>
<td>Tolerance</td>
<td>Irish (474 families)</td>
</tr>
<tr>
<td></td>
<td>Withdrawal</td>
<td>Irish (474 families)</td>
</tr>
<tr>
<td>Ma, 2003</td>
<td>Alcohol Consumption</td>
<td>Framingham Heart Study (329 families)</td>
</tr>
<tr>
<td>Prescott, 2006</td>
<td>AD symptoms</td>
<td>Irish (474 families)</td>
</tr>
<tr>
<td>Saccone, 2000</td>
<td>Max. alcoholic drinks</td>
<td>COGA (370 families, 2263 sibpairs)</td>
</tr>
<tr>
<td>Schuckit, 2001</td>
<td>LR to alcohol</td>
<td>COGA (745 ppl)</td>
</tr>
<tr>
<td>Schuckit, 2005</td>
<td>LR to alcohol</td>
<td>COGA (238 sibpairs)</td>
</tr>
<tr>
<td>Webb, 2011</td>
<td>LR to alcohol</td>
<td>Connecticut (238 sibpairs)</td>
</tr>
<tr>
<td>Wilhelmson, 2003</td>
<td>LR to alcohol</td>
<td>California (139 sibpairs)</td>
</tr>
<tr>
<td>Zhu, 2005</td>
<td>ALDX1 affected</td>
<td>COGA (119 families)</td>
</tr>
</tbody>
</table>
Linkage studies have also examined a range of quantitative traits related to alcoholism including consumption (Ma, Zhang et al. 2003), maximum number of drinks in a twenty-four hour period (Saccone, Kwon et al. 2000, Bergen, Yang et al. 2003, Kuo, Neale et al. 2006), alcohol craving (Ehlers, Wilhelmsen 2005), initial sensitivity (Kuo, Neale et al. 2006, Ehlers, Gizer et al. 2010), level of response to alcohol (Wilhelmsen, Schuckit et al. 2003, Schuckit, Wilhelmsen et al. 2005, Webb, Lind et al. 2011, Schuckit, Edenberg et al. 2001), severe drinking (Foroud, Bucholz et al. 1998, Gizer, Ehlers et al. 2011), age of AD onset (Kuo, Neale et al. 2006), alcohol withdrawal (Kuo, Neale et al. 2006) and a quantitative alcohol-related phenotype (Dick, Nurnberger et al. 2002).

Overall, there is little consistency between studies and linkage signals are weak, in keeping with results from other complex traits. The inability of linkage analyses to locate genetic risk loci for AD is likely due to insufficient samples sizes to detect numerous, small effects and also the heterogeneous nature of the AD phenotype. Despite these difficulties, the IASPSAD linkage study of AD-related phenotypes and AD symptom count identified two chromosomal regions with substantial overlap with previous studies. The first region, on chromosome 1, was linked to initial sensitivity and tolerance and will be discussed in detail in Chapter Four. The second region, chromosome 4q22–q32, was found in a linkage study of AD symptom count and had a maximum LOD score of 4.59 (Prescott, Sullivan et al. 2006). Seven other linkage studies (from three samples) of AD and related phenotypes have found evidence of linkage to the same chromosomal region. The first sample, comprised of Southwest Indian tribes, found linkage with AD (Long, Knowler et al. 1998) while the second, comprised of Mission Indian tribes, found linkage with AD severity (Ehlers, Gilder et al.
2004). The third, and largest, sample is COGA. At least partial overlap of linkage signals was found with AD (Williams, Begleiter et al. 1999), AD symptom count (Reich, Edenberg et al. 1998), AD severity (Corbett, Saccone et al. 2005), maximum drinks (Saccone, Kwon et al. 2000) and an anxiety/AD clinical phenotype (Dick, Nurnberger et al. 2002). Among the plausible candidate genes in this region is the ADH gene cluster, a well-recognized candidate gene cluster that has been frequently targeted in candidate gene association studies.

Candidate Gene Association Studies

As reported by Risch and Merikangas, association studies have superior power to detect genes underlying complex traits than linkage studies (Risch, Merikangas 1996). In contrast to linkage studies, association studies typically look within populations, not families, and currently focus on single nucleotide polymorphisms (SNPs). Association studies for alcohol dependence have focused on the obvious metabolic and neurotransmitter candidate genes. Due to the magnitude of literature related to this area, a brief overview of the most replicable and biologically relevant literature is included below.

Candidate Genes - Alcohol metabolizing enzymes

Ethanol is metabolized by three main enzymes (and obvious candidate genes for AD): alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH) and cytochrome P450 (CYP2E1). In the cytoplasm, ADH catalyzes the reaction whereby ethanol is oxidized, producing the toxic compound, acetaldehyde (Berg, Tymoczko et al. 2002).
ALDH then catalyzes the detoxification/oxidation of acetaldehyde in the mitochondria (Agarwal 2001). CYP2E1 encodes the major enzyme in an alternative pathway for ethanol metabolism in microsomes (Agarwal 2001). A brief discussion of the most relevant variants of each gene is included below.

Seven genes code for human ADH: ADH1A (formerly ADH1), ADH1B (formerly ADH2), ADHIC (formerly ADH3), and ADH4-7. They cluster on chromosome 4q and are divided into five classes based on their protein sequence and rate of metabolism (Edenberg 2007). Class 1 ADH enzymes, the primary ADH enzymes in the liver (Berg, Tymoczko et al. 2002), are dimers formed from combinations of α, β and γ polypeptides (encoded by the loci ADH1A, ADH1B and ADH1C, respectively) (Agarwal 2001). Of the three variants of ADH1B, ADH1B*2 and ADH1B*3 are considered “protective”, having much higher rates of ethanol metabolism than ADH1B*1 (Edenberg, Xuei et al. 2006). In a study of Native American Mission Indians with AD, there was increased frequency of the ADH1B*1 variant (Wall, Carr et al. 2003). The higher rate of metabolism seen with the *B2 and *B3 alleles increases the production of the toxin acetaldehyde (Osier, Pakstis et al. 1999) and is considered to be aversive to further drinking (Bierut, Goate et al. 2012). ADH1B*2 is found with greater frequency in Asian cultures (Shen, Fan et al. 1997) but has also been reported as protective in white Europeans (Borras, Coutelle et al. 2000), African and European Americans (Bierut, Goate et al. 2012) and an Australian twin sample (Macgregor, Lind et al. 2009). ADH1B*3 is associated with decreased risk for AD in Afro-Trinidadians (Moore, Montane-Jaime et al. 2007).

ALDH has two human isozymes: ALDH1 and ALDH2. While ALDH1 resides in the cytoplasm, variation exists in ALDH2, the fast-acting mitochondrial isozyme of ALDH
ALDH2 is encoded on human chromosome 12 and the single nucleotide transition that encodes the ALDH2*2 variant ablates enzymatic activity (Higuchi, Matsushita et al. 2004). This allele is also prevalent in Asian populations and results in a flushing reaction and aversion to alcohol (Gemma, Vichi et al. 2006, Li, Zhao et al. 2012). Studies have shown this allelic variant reduces the risk of AD in non-Asian populations as well (Luo, Kranzler et al. 2006) while other studies have shown the variant has no association with AD (Dickson, James et al. 2006).

Typically functioning after heavy drinking has saturated ADH, CYP2E1 also metabolizes ethanol (Gemma, Vichi et al. 2006). The CYP2E1 gene has thirteen polymorphisms: CYP2E1*1A-D, CYP2E1*2-4, CYP2E1*5A-B, CYP2E1*6 and CYP2E1*7C (Agarwal 2001). Several of the variants show association with AD or related phenotypes. CYP2E1*1B was reported to be more frequent in Mexican Indian alcoholics compared to controls (Montano Loza, Ramirez Iglesias et al. 2006). CYP2E1*6 was associated with AD in a Japanese sample (Iwahashi, Ameno et al. 1998). The CYP2E1*1D allele results in greater enzyme activity after alcohol consumption (McCarver, Byun et al. 1998) and may contribute to the risk for alcohol and nicotine dependence (Gemma, Vichi et al. 2006). A combined linkage and association study found significant association between level of response to alcohol and the CYP2E1*5B allele (Webb, Lind et al. 2011, Schuckit, Wilhelmsen et al. 2005, Wilhelmsen, Schuckit et al. 2003). However, multiple studies have found no association between CYP2E1 variants and AD or related phenotypes (e.g. (Plee-Gautier, Foresto et al. 2001, Carr, Yi et al. 1996)). While representing an obvious biological mechanism for involvement in AD, and providing strong association results (in particular in Asian
populations), variants in ethanol metabolizing enzymes fail to explain a large portion of the variance in AD in most populations.

**Candidate Genes - Gamma-Aminobutyric Acid**

Many behavioral consequences of alcohol use have been linked to the brain’s main inhibitory neurotransmitter, gamma-Aminobutyric acid (GABA). There are two GABA receptor classes, GABA\(_A\) and GABA\(_B\), with the former being implicated in behavioral responses to ethanol (Siegel, Agranoff et al. 1999, Kumar, Porcu et al. 2009). Mammalian GABA\(_A\) receptors are heteropentamers formed from 19 possible subunits: \(\alpha_{1-6}, \beta_{1-4}, \gamma_{1-4}, \delta, \epsilon\) and \(\rho_{1-3}\) (Siegel, Agranoff et al. 1999, Barnard, Skolnick et al. 1998). GABRA1, GABRA6, GABRB2, and GABRG2 cluster on human chromosome 5q (Dick, Edenberg et al. 2005). Multiple groups have found AD to be associated with genes in this cluster (Luo, Kranzler et al. 2006, Sander, Ball et al. 1999) yet other groups find no associations with AD (Dick, Edenberg et al. 2005, Dick, Plunkett et al. 2006, Song, Koller et al. 2003). Overall, the literature provides conflicting results for associations between GABA and AD.

One exception is seen with GABRA2 variants, which are among the best replicated genes in the candidate gene literature. Positive association results between GABRA2 variants (human chromosome 4p) and AD are found in the COGA sample (Edenberg, Dick et al. 2004), a large case-control study of European Americans (Covault, Gelernter et al. 2004), Japanese men (Roh, Matsushita et al. 2011), German AD inpatients (Fehr, Sander et al. 2006, Soyka, Preuss et al. 2008) and a meta-analysis of 8 studies and 14 GABRA2 variants (Zintzaras 2012). Recently, Olfson and Bierut
compiled a list of fifty-four candidate genes for AD from the literature (including \textit{ADH1B}, \textit{ALDH2} and \textit{CYP2E1}) and looked for replication in the Study of Addiction: Genetics and Environment (SAGE) genome-wide association dataset. The only gene that was associated with AD was \textit{GABRA2} (Olfson, Bierut 2012).

\textit{Candidate Genes - Dopamine}

It is thought that the reward circuitry in the brain contributes to the etiology of addiction. The reward pathway and addiction will be discussed in Chapter Two. Briefly, alcohol increases the amount of dopamine (the major neurotransmitter in the reward pathway) in areas of the brain important in the learned anticipation of reward (Ron, Jurd 2005). For this reason, genes involved in dopamine neurotransmission have been studied as candidates for AD. There are five known human receptor subtypes for dopamine, D1-D5, located on chromosomes 5, 11, 3, 11 and 4, respectively (Siegel, Agranoff et al. 1999). D3 and D4 receptor subtypes are typically not associated with AD in the literature (Parsian, Chakraverty et al. 1997, Gorwood, Martres et al. 1995), although at least one study found \textit{DRD3} to be associated with AD in a Caucasian sample (Hack, Kalsi et al. 2011). The D2 receptor subtype has been extensively investigated but the literature remains complicated. For example, Goldman et al found association with the D2 receptor subtype (Goldman, Urbanek et al. 1997), but the results failed to replicate in two subsequent studies (Gelernter, Kranzler 1999, Lee, Lu et al. 1999). The dopamine transporter gene, \textit{DAT1}, has also been frequently studied. The most investigated polymorphism is a variable number of tandem repeats in the 3'UTR of the gene. As reviewed in van der Zwaluw et al, most studies have found little
association with AD, but some association with alcohol withdrawal (van der Zwaluw, Engels et al. 2009). Other positive results within DAT1 include associations between rs6350, an exonic SNP, and both alcohol consumption (Lind, Eriksson et al. 2009) and AD (Hack, Kalsi et al. 2011).

As is exemplified in the association findings with dopamine and related genes, the candidate gene literature for alcoholism is full of false positives, conflicting reports, and sample-limited findings. In order to elucidate the genetic risk loci for AD, refined or new approaches are needed. Options for uncovering the missing heritability include better methods for candidate gene selection (to be covered in Chapter Two) and an unbiased method for locating new candidates.

**Genome-wide Association Studies**

The advantage of linkage studies is that they allow for an unbiased, genome-wide search for candidate genes. Association studies have superior power to detect genes for complex traits and map candidate genomic regions more tightly than linkage studies. Recently, advances in genotyping technology enabled genome-wide association studies (GWAS). Using GWAS, researchers can, without bias, identify fine-mapped areas of the genome associated with complex traits. GWAS involve genotyping individuals at SNPs, single base pair changes that occur in DNA, throughout the genome. Large-scale sequencing efforts such as The Human Genome project (Lander, Linton et al. 2001) and 1000 Genomes Project (1000 Genomes Project Consortium 2010) have contributed to the over 30 million validated SNPs in the latest dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/). SNPs can be in linkage
disequilibrium (LD) with other SNPs, meaning that two SNPs are associated more often than would be expected by chance alone. Two SNPs in high LD carry information about each other; that is by knowing information about one SNP you can predict information about the second SNP. The International Hapmap project has genotyped individuals from distinct ancestries (such as Northern and Western European or Han Chinese) allowing researchers to examine LD in each population (International HapMap Consortium 2003). By examining LD and haplotypes, or groups of alleles that are transmitted together, one can reduce the number of SNPs necessary to explain maximum information about the genome. Current GWAS arrays genotype approximately one million SNPs per individual.

To date, there have been eleven published GWAS which examine AD, two GWAS of AD symptoms and one small-scale meta-analysis. Six of those studies reported genome-wide significant findings after correction for multiple testing. Treutlein et al performed a GWAS in a German case-control sample and found significant association with two SNPs on chromosome 2: rs7590720 (intronic to PECR) and rs1344692 (intronic to AK093362) (Treutlein, Cichon et al. 2009). A second German case-control GWAS reported association with rs1789891, a SNP in LD with ADH1C (Frank, Cichon et al. 2012). The third study, comprised of predominately Caucasians, found several SNPs associated within KIAA0040 (Zuo, Zhang et al. 2012). Fourth, a small-scale meta-analysis by Wang and colleagues also found significant association in KIAA0040 and a second gene, THSD7B (Wang, Liu et al. 2011). The fifth study tested for association in the COGA sample and an Australian sample and found multiple SNPs associated within DSCAML1 (Wang, Liu et al. 2011). Finally, SNPs within C15orf53
were significantly associated with AD symptom count in the COGA families (Wang, Foroud et al. 2012). The remaining eight GWAS reported no genome-wide significant findings (Bierut, Agrawal et al. 2010, Edenberg, Koller et al. 2010, Heath, Whitfield et al. 2011, Lind, Macgregor et al. 2010, Zuo, Zhang et al. 2012, Zuo, Zhang et al. 2011, Kendler, Kalsi et al. 2011, Dickson, James et al. 2006). Within the AD GWAS literature, PKNOX2 has shown some evidence for replication. In the SAGE GWAS, the top SNP, though not genome-wide significant, was intronic to PKNOX2 (Bierut, Agrawal et al. 2010). Wang et al used a subset of the Australian twin-family study of alcohol use disorder (OZALC) GWAS sample and also found significant, though not genome-wide, association with SNPs within the gene (Wang, Liu et al. 2011).

In addition to GWAS of AD, multiple GWAS reports of alcohol-related phenotypes have been published. These are summarized in Table 1.2. Of those, there are three reports of genome-wide significant SNPs, all associated with alcohol consumption. Takeuchi and colleagues found association with rs671 (in ALDH2) in a Japanese cohort (Takeuchi, Isono et al. 2011). Looking within a Korean sample, Baik et al found association in SNPs within 6 genes; c12orf51, CCDC63, MYL2, OAS3, CUX2, and RFH3A (Baik, Cho et al. 2011). SNPs within c12orf51 were also in LD with ALDH2. Lastly, a large, predominately Caucasian GWAS reported significant association of rs6943555 in AUTS2 (Schumann, Coin et al. 2011). The remaining GWAS presented in Table 2 reported no genome-wide significant results (Agrawal, Wetherill et al. 2012, Chen, Xiong et al. 2012, Heath, Whitfield et al. 2011, Wang, Liu et al. 2012). There were no replicated SNPs or loci among the top results in these studies.
Table 1.2. Survey of Published GWAS for AD-related Phenotypes.

<table>
<thead>
<tr>
<th>Study</th>
<th>Phenotype</th>
<th>Study Type</th>
<th>Primary sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baik, 2011</td>
<td>Alcohol consumption</td>
<td>Population</td>
<td>1721 males</td>
</tr>
<tr>
<td>Schuman, 2011</td>
<td>Alcohol consumption</td>
<td>Population</td>
<td>26,316 individuals</td>
</tr>
<tr>
<td>Takeuchi, 2011</td>
<td>Alcohol consumption</td>
<td>Population</td>
<td>1462 individuals</td>
</tr>
<tr>
<td>Agrawal, 2012</td>
<td>Alcohol craving</td>
<td>Ascertained</td>
<td>1899 AD cases, 1938 controls</td>
</tr>
<tr>
<td>Chen, 2012</td>
<td>Alcohol drinking</td>
<td>Population</td>
<td>904 individuals</td>
</tr>
<tr>
<td>Heath, 2011</td>
<td>Alcohol drinking</td>
<td>Ascertained</td>
<td>6852 individuals</td>
</tr>
<tr>
<td>Wang, 2012</td>
<td>Alcohol withdrawal symptoms</td>
<td>Ascertained</td>
<td>461 cases, 408 controls</td>
</tr>
</tbody>
</table>

Refined GWAS methodology

A review of the GWAS literature for AD and related phenotypes shows a minimal number of genome-wide significant results and few replicated loci. Association studies are based on the common disease - common variant (CDCV) hypothesis; that there are a large number of loci of small effect size that contribute to risk for AD (Rowe, Tenesa 2012). Threshold approaches looking only at SNPs that pass an a priori significance level may leave researchers vulnerable to reporting false positives while also missing true association signals.

Refined GWAS methodology - FDR control

One simple technique to explore a larger number of variants is the false discovery rate (FDR) procedure. The FDR approach attempts to limit the number of
false negatives while allowing for a few false positives, attempting to ensure that truly significant results are not overlooked (Benjamini, Hochberg 1995). Within each dataset, a FDR value (q-value) is empirically set as a cutoff point. A q-value of 0.2, for example, is interpreted as 20% of the p-values at that level or more significant representing false positives. Therefore, p-values whose associated q-values are more significant than the appointed cutoff are statistically more likely to be true risk variants (Neale, Ferreira et al. 2008). Using a FDR control allows for a larger set of results to be analyzed, which may represent true association signals lost in the noise of large association studies.

As proof of principle, a FDR approach was used in a recent, unpublished GWAS of AD in the IASPSAD sample with positive results (Hack et al, unpublished). Only two SNPs passed the *a priori* significance threshold of 3.06 x 10^-8. However, using a FDR q-value < 0.2, twenty-six other loci were identified as associated with AD. Mutations in at least two of the loci have previous associations with alcohol-related phenotypes in model organisms: *RYR3* and *KLF2*. These results would have been missed by just considering the top, genome-wide significant results.

**Refined GWAS methodology – Gene-based association**

A second technique to avoid the threshold approach is to use gene-based or pathway analyses. Gene-based association studies begin by assigning SNPs to genes and then typically use either the minimum p-value per gene (e.g., (Bigdeli, Maher et al. 2011)) or a combination of all p-values (e.g., using the Simes procedure (Li, Gui et al. 2011) or Fisher’s combination test (Curtis, Vine et al. 2008)). In Bigdeli et al and Li et al, the gene-based approach identified more significantly associated variants than
individual SNP association tests. Using one SNP per gene cannot test if modest associations in multiple variants within a gene are significantly related to the phenotype of interest. Techniques using multiple SNPs per gene are an improvement, however they do not take into account LD. If multiple associated SNPs are in LD, using all of them will artificially inflate the test statistic.

The Versatile Gene-Based Test for GWAS (VEGAS) assigns SNPs to genes, accounts for LD through a series of Monte Carlo simulations and calculates an empirical gene-based test statistic for each gene (Liu, McRae et al. 2010). To date, this is the best option for gene-based association as it combines the use of multiple variants per gene while also taking into account the LD between SNPs. Unpublished results from a GWAS of maximum alcohol consumption in the IASPSAD sample (Chapter Five) identified a gene using VEGAS with multiple previous associations with addiction, alcohol consumption and neuroadaptation following chronic alcohol use. No single SNPs from this candidate gene were within the top results.

Refined GWAS methodology –Pathway analyses

Pathway analyses look for modest associations, in aggregate, among genes in a predefined pathway. These methods allow researchers to look for overrepresentation of association signals in either large groups of genes forming a functional pathway or in a hand-picked gene list. A disadvantage, compared to the gene-based association analyses, is that these methods rely on pre-defined gene pathways. The researcher is testing GWAS data for groups of genes already known to be functionally related.
Uncharacterized loci that might contribute to a phenotype of interest would therefore be missed.

One popular pathway analysis method is Association LIst Go AnnoTatOR (ALIGATOR) which examines SNP association in gene sets defined by Gene Ontology (GO) and tests whether the number of genes significantly associated in each set is greater than the number expected by chance (Holmans, Green et al. 2009). GO is a database that collects and organizes information regarding gene products’ biological, cellular and molecular aspects (Ashburner, Ball et al. 2000). A recent GWAS of information processing speed found no genome-wide significant associations with single SNPs but several biological pathways associated with related traits after using ALIGATOR (Luciano, Hansell et al. 2011).

Other pathway analyses can input a user-defined set of genes and test for association of SNPs, in aggregate, in these genes compared to a random gene list. Gene Set Enrichment Analysis (GSEA), originally developed for microarrays (Subramanian, Tamayo et al. 2005), has been modified for GWAS data (Wang, Li et al. 2007). SNPs are assigned to genes, each gene is represented by the minimum p-value, and an enrichment score is calculated by comparing the number of positive associations in the predefined gene set with a random gene set. FDR, gene-based and pathway-based analyses allow researchers to examine their data in ways not possible with single SNP association. Oftentimes these methods reveal new candidate genes or pathways for phenotypes of interest.
Part II: A brief overview of the phenotypic complexity of AD

As discussed previously, alcohol dependence (AD) is a dichotomous, broad and clinical diagnosis. Two individuals can meet diagnostic criteria for AD with different symptoms. Furthermore, there are many different risk pathways for AD. Pathways to AD include physiological traits that affect how an individual processes and reacts to ethanol. They also include co-occurring or predisposing psychiatric conditions such as depression. Examining AD as a phenotype is useful to test variants for risk to a broader liability for alcohol use problems. However, breaking the phenotype down into individual components of risk is a powerful methodology. Two risk pathways for AD are explored in this dissertation: physiological alcohol-related phenotypes and internalizing psychopathology.

Alcohol-related phenotypes

Researchers have had modest success finding genes associated with AD, but analyzing alcohol-related phenotypes (ARPs) is an attractive alternative. ARPs are specific factors thought to contribute to the overall risk for AD. While AD is likely due in part to the action of numerous genes, ARPs, like physiological measurements, may be closer to the underlying genetic risk factors, may be localized to specific regions of the genome and are more likely to comprise smaller gene networks. The linkage and GWAS sections of this Introduction have presented results from both the AD analyses and those of ARPs such as level of response to alcohol, alcohol consumption, tolerance and withdrawal. Chapters Four and Five present an in-depth review of two ARPs.
Comorbidity

Comorbidity is the norm rather than the exception in common psychiatric (e.g. major depression) and substance use disorders (e.g. alcohol dependence)(Kessler, Nelson et al. 1996, Regier, Farmer et al. 1990). Internalizing disorders, characterized by negative mood states, include major depression, dysthymia, generalized anxiety disorder and phobias. Externalizing disorders, characterized by acting out and behavioral disinhibition, include substance use disorders, adult antisocial personality disorder and conduct disorder. AD correlates most strongly with the externalizing disorders, but is significantly correlated with internalizing disorders as well (Kessler, Nelson et al. 1996, Kessler, Crum et al. 1997, Kessler, Chiu et al. 2005, Regier, Farmer et al. 1990, Dawson, Goldstein et al. 2010). In concert with ARPs, it is only by understanding the relationship between AD and comorbid phenotypes, and the genetic architecture contributing to each, that researchers can begin to untangle the genetics of alcohol use disorders. Chapter Six discusses these concepts in detail and presents ongoing work in examining the relationship between AD and comorbid internalizing disorders.

Conclusion

A strong foundation of evidence supports the existence of genetic risk factors for AD and related phenotypes. Past and current methodologies have uncovered few replicable genetic risk loci for AD and AD-related phenotypes. This dissertation examines four different methods to uncover genetic risk factors for Alcohol Dependence. In Chapter Two, the model organism literature is used to provide a better
candidate gene selection method for alcohol-related traits in humans. Chapter Three focuses on an early GWAS methodology to identify genes of interest and then subsequent analyses of the top results. The final three chapters represent analyses in the genomics era of alcohol research. Chapters Four and Five present results from GWAS studies of two quantitative traits known to influence risk for AD; initial sensitivity and maximum alcohol consumption in twenty-four hours. In Chapter Six, preliminary results from the creation of a multivariate AD/internalizing phenotype are presented in an effort to identify risk loci for a broader comorbid disease liability in GWAS data.
Investigating the Role of \textit{GPSM1} in Human Alcohol Dependence and Craving Behaviors

Introduction

Among the many factors contributing to the high prevalence of Alcohol Dependence (AD) is the frequent endorsement of craving among high risk alcohol drinkers. Even among individuals who undergo treatment, up to 60\% will relapse within a year (Hunt, Barnett et al. 1971). Factor analyses consistently show craving loading highly onto the existing criteria for Alcohol Dependence (AD) as defined by the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (Cherpitel, Borges et al. 2010, Keyes, Krueger et al. 2011, Casey, Adamson et al. 2012, Agrawal, Wetherill et al. 2012). In part due to this evidence, the upcoming DSM-V will include craving in the AD diagnosis, defined as the ‘strong desire or urge to use’ alcohol (dsm5.org).

\textit{The biology of craving}

To discuss the proposed biology of craving, it is necessary to briefly introduce both Koob’s theory of addiction and the brain’s reward system. A prominent theory developed by Koob explains addiction as hedonic homeostatic dysregulation (Koob, Le Moal 1997). He describes addiction as a cycle of dysregulation in the brain’s reward system. In the case of alcohol addiction, individuals would initially drink to receive pleasurable effects but eventually end up drinking to avoid the negative consequences associated with withdrawal. This was termed counteradaptation and is based on the
opponent-process theory (Solomon 1980). The other main tenant of Koob’s theory is sensitization. Sensitization, seen as the increasing effect of a drug of abuse after repeated use, results in a hypersensitive brain reward system (Koob, Le Moal 1997). Counteradaptation and sensitization therefore together contribute to drug-seeking and eventual addiction (Koob, Le Moal 1997). Changes in the brain after repeated drug exposure, or imbalances during withdrawal, are loosely termed neuroadaptations (Anton 1999).

Alcohol increases the amount of extracellular dopamine in the mesolimbic dopaminergic pathway, the main component of the brain’s reward system (Ron, Jurd 2005). Dopaminergic neurons project from the ventral tegmental area to the nucleus accumbens, amygdala and other regions of the basal forebrain (Nestler 2001). Aside from the nucleus accumbens, a region implicated in the learned anticipation of reward, the basolateral amygdala and extended amygdala are integral in reward processing (Koob 2006). The basolateral amygdala is thought to contribute to relapse by its integration of drug-related memories and stress (Belujon, Grace 2011), resulting in cue-induced craving (Koob 2006). Part of the extended amygdala, the central nucleus of the amygdala is involved in drug-reinforcement, integrating stimuli as positive (incentive salience) or negative (aversive salience) (Koob 2006, Ode, Winters et al. 2012). As mentioned earlier, Koob’s model of addiction suggests that sensitization results in hypersensitive mesolimbic dopamine system and this is thought to manifest itself as craving (Anton 1999).
The genetics of craving

The heritability of craving has been estimated to be 65% (Ehlers, Wilhelmsen 2005). Multiple genetic risk loci for craving have been reported in the literature including SRD5A2 (Lenz, Schopp et al. 2012), AR and CYP19A1 (Lenz, Jacob et al. 2009, Lenz, Heberlein et al. 2011), NK1R (Seneviratne, Ait-Daoud et al. 2009), SCNA (Bonsch, Greifenberg et al. 2005, Bonsch, Reulbach et al. 2004, Foroud, Wetherill et al. 2007), 5-HTTLPR (Bleich, Bonsch et al. 2007), GABRA6 (Han, Bolo et al. 2008), TACR1 (Blaine, Claus et al. 2012) and dopamine D2 receptors (Heinz, Siessmeier et al. 2005). Recently, Agrawal et al performed a genomewide association study (GWAS) of craving (Agrawal, Wetherill et al. 2012). No single nucleotide polymorphisms (SNPs) were genomewide significant. The top hit, rs2454908, was located on chromosome 7 within ITGAD.

Model organism studies of addiction and withdrawal have implicated GPSM1, or AGS3, and drug-seeking behavior. GPSM1, G protein signaling modulator 1, encodes a guanine dissociation inhibitor located on the long arm of human chromosome 9 (9q34.3). The rodent ortholog of GPSM1, Ags3, has been previously implicated in morphine withdrawal (Fan, Jiang et al. 2009) and cocaine- (Bowers, McFarland et al. 2004) and heroin- (Yao, McFarland et al. 2005) seeking during withdrawal. Of particular interest, Bowers and colleagues investigated differential protein expression of Ags3 and ethanol-seeking in rats (Bowers, Hopf et al. 2008). Briefly, rats taught to self-administer ethanol showed correlated increases in both ethanol-seeking behavior and expression of Ags3 in the nucleus accumbens after 3 weeks of ethanol withdrawal. RNAi knockdown of Ags3 diminished ethanol-seeking behavior. This increase in Ags3
expression seen in the rodent model during withdrawal may represent a neuroadaptation which occurs during periods of abstinence after heavy alcohol use.

This chapter presents work investigating the role of $GPSM1$ in humans. SNPs within $GPSM1$ were genotyped in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) sample and tested for association with AD and craving related behaviors. Replication was sought in another ascertained sample, the Collaborative Study on the Genetics of Alcoholism (COGA). Finally, $GPSM1$ messenger RNA (mRNA) expression differences were tested in human chronic alcoholic and control post-mortem brain tissues.

**Methods**

**Primary sample: Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD)**

A detailed description of the sample is provided elsewhere (Prescott, Sullivan et al. 2006). Briefly, probands were ascertained in clinical treatment facilities and hospitals in the Republic of Ireland and Northern Ireland between 1998 and 2002. Probands were interviewed using an adapted version of the Semi-Structured Assessment of the Genetics of Alcoholism (SSAGA) and diagnosed with AD using DSM-IV criteria. The sample is ethnically homogenous; a stipulation for inclusion in the study required each proband have all 4 grandparents born in Ireland, Northern Ireland, Scotland, Wales or England. Controls were recruited from the National Guard and army reserve in the Republic of Ireland and from volunteers donating at the Northern Ireland Blood Transfusion Service in Northern Ireland. Controls had no reported history of alcoholism.
From the original linkage sample, 562 genetically independent AD cases and 569 controls were selected for genotyping analyses.

Replication sample: Collaborative Study on the Genetics of Alcoholism (COGA)

Probands for the COGA sample, details of which are described elsewhere (Dick, Plunkett et al. 2006), were recruited throughout the United States in 6 geographic locations from alcohol treatment programs. Large families were preferentially ascertained with a proband required to have 2 or more family members in the COGA sampling area and preferably sibships greater than 3. All individuals were given the SSAGA interview and AD diagnosed using DSM-IV criteria. From the larger COGA sample, 847 genetically independent AD cases and 552 controls of European ancestry were used.

Phenotypes

Three phenotypes were analyzed in both samples: AD (DSM-IV), craving and a sum score. Craving was extracted from the SSAGA interview ("Do you have a strong desire to drink when you cannot?") as a dichotomous variable. The sum score phenotype was calculated by totaling how many of the following phenotypes a subject endorsed: craving (as above), DSM-IV AD Symptom 3 ("Have you started drinking when you promised yourself you would not or have you ever drunk more than you intended?") and DSM-IV AD Symptom 5 ("Have you ever spent so much time drinking alcohol or recovering from it that you had little time for anything else?"). Both DSM-IV Symptom 3 and 5 were also extracted from the SSAGA interview as dichotomous variables.
**SNPs and genotyping: IASPSAD**

SNPs were genotyped using monoplex genotyping. Monoplex reactions were performed using Taqman Assays-on-Demand (Applied Biosystems, Foster City, CA, USA). One SNP, rs28539249, was genotyped using a custom Taqman Assay (sequence available upon request). Genotypes were called using the Analyst AD fluorescence detector (LJL Biosystems, Sunnyvale, CA, USA) and automated Excel template (van den Oord, Jiang et al. 2003). To minimize technical variability, all reaction steps were performed using the Eppendorf 5075 automated liquid handler. Stringent evaluation of initial data is important to avoid artificial effects of genotyping errors; therefore, all genotypes were independently assessed by two raters. Ambiguous calls were discussed and in cases of non-resolution, genotypes were dropped from the analyses.

**SNPs and genotyping: COGA**

Genotypes for analysis were extracted from the COGA GWAS dataset. As detailed elsewhere (Edenberg, Koller et al. 2010), COGA samples were genotyped on Illumina HumanHap1M Bead Chips (Illumina, San Diego, CA) by the Center for Inherited Disease Research (CIDR). After stringent quality control measures, 853,375 SNPs were available for analysis. The dataset was subsequently imputed with IMPUTE2 (Howie, Donnelly et al. 2009) using the March 2012 version (v3) of the 1000 Genomes reference panel (1000 Genomes Project Consortium 2010). SNPs were filtered for minor allele frequency (MAF) <1% and information content ≤3%. After filtering, 8.5 million SNPs remained.
Data analysis

IASPSAD single marker association analyses and Hardy-Weinberg Equilibrium (HWE) testing were performed using Plink v1.07 (http://pngu.mgh.harvard.edu/purcell/plink) (Purcell, Neale et al. 2007). Logistic regression was performed on the IASPSAD case-control sample, with sex as a covariate. Linear regression was performed on the AD cases in the IASPSAD sample, with sex and age as covariates. SNPSpD (Nyholt 2004) was used to determine the number of independent of tests for multiple testing corrections. Analysis with SNPSpD in the IASPSAD sample identified 5 independent SNPs, yielding an adjusted threshold of p<0.0091. COGA single marker association analyses were performed using Plink v1.07 and the imputed dosages. Within the case-control sample, SNPs were tested for association using sex as a covariate. Craving and the sum score phenotypes were tested within cases using age and sex as covariates. Phenotypic distributions are shown in Tables 2.1 and 2.2.

<table>
<thead>
<tr>
<th>Table 2.1. Phenotype distributions in the case-control analyses.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotype</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>IASPSAD</td>
</tr>
<tr>
<td>AD</td>
</tr>
<tr>
<td>Craving</td>
</tr>
<tr>
<td>COGA</td>
</tr>
<tr>
<td>AD</td>
</tr>
<tr>
<td>Craving</td>
</tr>
</tbody>
</table>
Table 2.2 Phenotype distributions in the case-only analyses.

<table>
<thead>
<tr>
<th></th>
<th>Craving</th>
<th>Sum Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>IASPSAD</td>
<td>428</td>
<td>107</td>
</tr>
<tr>
<td>COGA</td>
<td>486</td>
<td>360</td>
</tr>
</tbody>
</table>

Gene expression study: sample

Superior frontal cortex (CTX), central nucleus of amygdala (CNA) and basolateral nucleus of amygdala (BLA) tissues from human chronic alcoholics and controls from the New South Wales Tissue Resource Centre (NSW TRC) located at the University of Sydney were used in this study. As described in Ponomarev, 2012, DSM-IV diagnosis of cases was based on clinical review. A GWAS was performed on 41 chronic alcoholics and 41 cases from the NSW TRC using Affymetrix 6.0 arrays (Riley et al, unpublished results). Expression data was available for 17 chronic alcoholics and 15 controls from the NSW TRC from a genome-wide expression study with Illumina (Illumina, San Diego, CA) HumanHT-12 whole genome expression bead chips (Ponomarev, Wang et al. 2012). The samples from these studies were not congruent, so the overlapping samples between the datasets, 12 chronic alcoholics and 8 controls, were chosen for analyses.
Gene expression study: data analysis

Sex, age, PMI (post-mortem interval), liver pathology, neuropathology, toxicology, brain pH, and smoking demographics were provided by the NSW TRC. As 1 out of the 20 samples was female, sex was not included as a covariate. RNA integrity number (RIN) values were provided for each tissue, however a Fisher’s exact test was non-significant for differences between cases and controls in the CTX, CNA and BLA (p=0.586, 1, and 0.906, respectively). RIN was therefore not included as a covariate. Liver pathology, neuropathology and toxicology were re-coded numerically for subsequent analyses. Principal components analysis (PCA) was run in SASv9.3 using age, PMI, liver pathology, neuropathology, toxicology, brain pH, and smoking. The first 5 components explained 90% of the variance in the dataset (Figure 2.1). Component scores for the first 5 components were used as covariates in the analyses.

Differences in **GPSM1** mRNA expression (Illumina identifier ILMN_1709307) between cases and controls were analyzed using the PROC GLM – MANOVA test in SAS. The GWAS dataset (Riley et al, unpublished results) was imputed with IMPUTE2 (Howie, Donnelly et al. 2009) using the March 2012 version (v3) of the 1000 Genomes reference panel (1000 Genomes Project Consortium 2010). Association tests for allele-specific expression differences were performed in Plink v1.07 with imputed dosages and using the aforementioned component scores as covariates.
Fig. 2.1. **Eigenvalues and scree plot results for principal components analysis.**
Age, PMI, liver pathology, neuropathology, toxicology, brain pH, and smoking demographics from 12 alcoholic cases and 8 controls were run in a PCA. The first five components explained 90% of the variance.

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.12699</td>
<td>0.30390</td>
<td>0.30390</td>
</tr>
<tr>
<td>2</td>
<td>1.42603</td>
<td>0.20370</td>
<td>0.50760</td>
</tr>
<tr>
<td>3</td>
<td>1.22804</td>
<td>0.17540</td>
<td>0.68300</td>
</tr>
<tr>
<td>4</td>
<td>0.88309</td>
<td>0.12620</td>
<td>0.80920</td>
</tr>
<tr>
<td>5</td>
<td>0.63580</td>
<td>0.09080</td>
<td>0.90000</td>
</tr>
<tr>
<td>6</td>
<td>0.44292</td>
<td>0.06330</td>
<td>0.96330</td>
</tr>
<tr>
<td>7</td>
<td>0.25714</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Results

Nine SNPs within *GPSM1* were genotyped in the IASPSAD sample. Two SNPs genotyped poorly and were dropped. No SNPs showed significant deviation from HWE ($p < 0.001$). One SNP was removed for excessive missingness (19%). Seventy-four samples were removed for having more than 2 missing genotypes (>33% missingness). After data cleaning, a total of 6 SNPs were analyzed for tests of association on a possible 535 AD cases and 522 controls.

One SNP, rs28626972, was nominally associated with AD in the IASPSAD case-control sample but did not pass multiple testing correction (Table 2.3). No SNPs were significantly associated in the craving cases versus controls analysis. Within AD cases, rs28439345 was nominally associated, only, with craving (Table 2.4). Rs28439345 and rs28536668 were nominally associated with the sum score phenotype and rs28439345 remained significantly associated after multiple testing correction.

Table 2.3. Association Results for *GPSM1* in the IASPSAD and COGA case-control samples.

<table>
<thead>
<tr>
<th>SNP</th>
<th>BP</th>
<th>IASPSAD</th>
<th>COGA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AD</td>
<td>Craving</td>
</tr>
<tr>
<td>rs28380074</td>
<td>139225139</td>
<td>0.3900</td>
<td>0.5652</td>
</tr>
<tr>
<td>rs28626972</td>
<td>139227544</td>
<td>0.0427</td>
<td>0.1106</td>
</tr>
<tr>
<td>rs28536668</td>
<td>139234512</td>
<td>0.2131</td>
<td>0.5547</td>
</tr>
<tr>
<td>rs28439345</td>
<td>139243790</td>
<td>0.1860</td>
<td>0.6425</td>
</tr>
<tr>
<td>rs3812547</td>
<td>139252495</td>
<td>0.2432</td>
<td>0.2720</td>
</tr>
<tr>
<td>rs1128905</td>
<td>139253839</td>
<td>0.1008</td>
<td>0.2639</td>
</tr>
</tbody>
</table>

Nominal p-values are shown for association between each of the six SNPs within *GPSM1* and two phenotypes: AD and craving.
Table 2.4. Association Results for GPSM1 within IASPSAD and COGA AD cases.

<table>
<thead>
<tr>
<th>SNP</th>
<th>BP</th>
<th>IASPSAD Craving</th>
<th>IASPSAD Sum Score</th>
<th>COGA Craving</th>
<th>COGA Sum Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs28380074</td>
<td>139225139</td>
<td>0.5866</td>
<td>0.6848</td>
<td>0.6754</td>
<td>0.1909</td>
</tr>
<tr>
<td>rs28626972</td>
<td>139227544</td>
<td>0.3098</td>
<td>0.0873</td>
<td>0.0080</td>
<td>0.6487</td>
</tr>
<tr>
<td>rs28536668</td>
<td>139234512</td>
<td>0.1696</td>
<td>0.0457</td>
<td>0.5022</td>
<td>0.3162</td>
</tr>
<tr>
<td>rs28439345</td>
<td>139243790</td>
<td>0.0188</td>
<td>0.0035</td>
<td>0.4459</td>
<td>0.0472</td>
</tr>
<tr>
<td>rs3812547</td>
<td>139252495</td>
<td>0.8304</td>
<td>0.7425</td>
<td>0.1853</td>
<td>0.3801</td>
</tr>
<tr>
<td>rs1128905</td>
<td>139253839</td>
<td>0.2583</td>
<td>0.1601</td>
<td>0.4853</td>
<td>0.7099</td>
</tr>
</tbody>
</table>

Nominal p-values are shown for association between each of the six SNPs within GPSM1 and two phenotypes: craving and a sum score calculated from endorsement of three possible craving-related behaviors.

Based on the association results in our sample, we sought to replicate our findings in an independent sample. Association results from the COGA case-control sample, and within the case-only analysis, are shown in Tables 2.3 and 2.4, respectively. No SNPs were associated in the COGA case-control sample. Within COGA cases, rs28626972 remained associated with craving after a Bonferroni correction for the six tested SNPs (p=0.048). Rs28439345, identified in the IASPSAD sample as associated with the sum score, was also significantly associated with the sum score phenotype within the COGA cases.

We then examined GPSM1 mRNA expression in human postmortem brain tissues from the NSW TRC. There were no expression differences between cases and controls (MANOVA p=0.1831). Allele specific expression differences (rs28439345) were tested in three brain regions: the basolateral nucleus of amygdala (BLA), central
nucleus of amygdala (CNA), and superior frontal cortex (CTX). No differences were found (p=0.716, 0.323 and 0.831, respectively).

Discussion
Relapse is a trademark of addiction. With all drugs of abuse, including alcohol, addicts are plagued by high rates of recidivism and difficulty remaining abstinent. The phenotype seen in the rat model described by Bowers et al seems to approximate the compulsive, alcohol-seeking behavior seen in many chronic alcoholics during periods of withdrawal. The report also implicated increased expression of GPSM1 in drug-seeking behavior during withdrawal. Therefore, GPSM1 was examined in human AD using association and expression studies.

A series of association analyses with AD and craving-related phenotypes was performed. First, the broad diagnostic phenotype of AD was studied to ascertain if variation in GPSM1 distinguished AD cases from controls. No SNPs survived multiple testing correction in the IASPSAD sample and no SNPs were nominally associated in the COGA sample. All subsequent association analyses involved more specific phenotypes. There was no association in the case-control analysis of craving, suggesting that variation in the gene did not distinguish craving AD cases from controls. It was then asked if variation in GPSM1 distinguished cases endorsing craving from cases who did not report craving alcohol. No SNPs survived multiple testing correction in the IASPSAD sample. Rs28626972 was significantly associated with craving within the COGA AD cases and remained significant after Bonferroni correction for the six tested SNPs. Finally, it was determined that variation in GPSM1 was associated with a
sum score phenotype comprised of three craving-related behaviors. Rs28439345 was associated with the sum score in the IASPSAD cases after multiple testing correction and was replicated in the COGA AD cases (p=0.047). In both samples, the minor allele of the SNP was associated with decreased sum score, or a decrease in the craving-related behaviors included in the analysis.

Rs28439345 is intronic to GPSM1. There are no known reported associations between this variant and alcohol-related phenotypes in the literature. To further investigate the role of GPSM1 and rs28439345 in human alcoholism, mRNA expression differences between chronic alcoholic and control post mortem amygdala and cortex tissues were examined. There were no GPSM1 expression differences in any of the tested brain regions. Similarly, no allele-specific expression differences were seen. Further research is necessary to determine the effect of variation within GPSM1 on the queried phenotypes.

There are three main limitations in this study. First, the experimental paradigms used in the model organism studies are impossible to replicate in human samples. Secondly, an assumption is made that the ethanol-seeking behavior reported in rats in the literature is related to the craving behaviors investigated in both human samples. Finally, Bowers et al found protein expression differences while this study only examined mRNA expression. It is possible that there are protein expression differences in our tissue samples while mRNA levels remain unchanged.

This chapter has presented results showing variation within GPSM1 is associated with a craving sum score phenotype in two independent, clinically ascertained samples. This provides a link between human craving of alcohol and a
gene implicated in drug-seeking behavior in rodents. Furthermore, this work suggests that the model organism literature provides a targeted, improved method for selection of candidate genes in addiction research.

**Future Directions**

Two additional studies are currently underway. The first is a gene set enrichment analysis (GSEA-SNP, see Introduction) of genes related to \textit{GPSM1} function. GSEA-SNP will detect if modest effect sizes in aggregate are associated with craving related behaviors in AD cases. A thorough literature search was done for genes with known interactions with \textit{GPSM1}. Along with genes whose biological function suggests involvement in the \textit{GPSM1} pathway, they were compiled into the gene set shown in Table 2.5.

### Table 2.5. Gene set of \textit{GPSM1}-related genes.

<table>
<thead>
<tr>
<th>ADCY1</th>
<th>ADRA2B</th>
<th>DRD4</th>
<th>GNB2</th>
<th>GNG5</th>
<th>GRM7</th>
<th>HTR5A</th>
<th>OSTM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADCY10</td>
<td>ADRA2C</td>
<td>DRD5</td>
<td>GNB3</td>
<td>GNG7</td>
<td>GRM8</td>
<td>HTR5A</td>
<td>PPYR1</td>
</tr>
<tr>
<td>ADCY2</td>
<td>ADRB1</td>
<td>FRMPD1</td>
<td>GNB4</td>
<td>GNG8</td>
<td>HRH2</td>
<td>HTR7</td>
<td>RGS4</td>
</tr>
<tr>
<td>ADCY3</td>
<td>ADRB2</td>
<td>GABBR1</td>
<td>GNB5</td>
<td>GNGT1</td>
<td>HRH3</td>
<td>INSC</td>
<td>RIC8A</td>
</tr>
<tr>
<td>ADCY4</td>
<td>ADRB3</td>
<td>GABBR2</td>
<td>GNG10</td>
<td>GNGT2</td>
<td>HRH4</td>
<td>NPY1R</td>
<td></td>
</tr>
<tr>
<td>ADCY5</td>
<td>CCDC158</td>
<td>GNAI1</td>
<td>GNG11</td>
<td>GPSM1</td>
<td>HTR1A</td>
<td>NPY2R</td>
<td></td>
</tr>
<tr>
<td>ADCY6</td>
<td>CHRM2</td>
<td>GNAI2</td>
<td>GNG12</td>
<td>GPSM2</td>
<td>HTR1B</td>
<td>NPY5R</td>
<td></td>
</tr>
<tr>
<td>ADCY7</td>
<td>CHRM4</td>
<td>GNAI3</td>
<td>GNG13</td>
<td>GRM2</td>
<td>HTR1D</td>
<td>OPRD1</td>
<td></td>
</tr>
<tr>
<td>ADCY8</td>
<td>DRD1</td>
<td>GNAL</td>
<td>GNG2</td>
<td>GRM3</td>
<td>HTR1E</td>
<td>OPRK1</td>
<td></td>
</tr>
<tr>
<td>ADCY9</td>
<td>DRD2</td>
<td>GNAS</td>
<td>GNG3</td>
<td>GRM4</td>
<td>HTR1F</td>
<td>OPRL1</td>
<td></td>
</tr>
<tr>
<td>ADRA2A</td>
<td>DRD3</td>
<td>GNB1</td>
<td>GNG4</td>
<td>GRM6</td>
<td>HTR4</td>
<td>OPRM1</td>
<td></td>
</tr>
</tbody>
</table>
Finally, copy number variation (CNV) is being assessed. There are multiple reports of structural variation in the region. Losses or deletions have been reported by six groups, a gain by one group and an insertion by another. Figure 2.2 summarizes the known variation in the region. Iafrate et al reported 1 loss in 39 control samples (ancestry information not provided) (Iafrate, Feuk et al. 2004) while Redon et al reported losses in 9 of 270 Hapmap samples, all of which had European ancestry (Redon, Ishikawa et al. 2006). Both studies used Comparative Genomic Hybridization (CGH). In 2008, Kidd et al, using a combination of sequencing, mapping and CGH, found 1 loss in a Hapmap sample of European descent (Kidd, Cooper et al. 2008). Jakobsson et al identified 5 losses in Hapmap individuals of Siberian and Pakistani descent using Illumina Bead Chip calls (Jakobsson, Scholz et al. 2008). Pang et al (Pang, MacDonald et al. 2010) reported a deletion in the region from the genome of Craig Venter (Levy, Sutton et al. 2007). The last group to report a loss in the region, Altshuler et al, reported 11 individuals with losses from the 1000 Genomes sequencing project (1000 Genomes Project Consortium 2010). Park et al reported a gain in an Asian individual using CGH and subsequent sequencing (Park, Kim et al. 2010) and Teague et al reported an insertion in a Hapmap individual of European descent using optical mapping (Teague, Waterman et al. 2010).

In summary, there are 32 reported losses or deletions and both a gain and insertion reported either within or proximal to *GPSM1*. Further evidence is found in the imputed data from an unpublished GWAS in the IASPSAD sample (see Chapter Four). In this region on chromosome 9, the imputed genotyping probability drops to below 7% (ImputedFail 1 in Table 2.2), and then hovers around or below 30% for another 13kb
(ImputedFail 2 in Table 2.2). This could be indicative of a structural rearrangement in the DNA interrupting sequencing and genotyping of this region.

Multiple studies have reported structural variation in the region of *GPSM1*, however several of the studies used the same sample (i.e. Hapmap individuals used by both Redon et al and Jakobsson et al) and did not find the same variants in the same individuals. Therefore, it is integral within the IASPSAD and NSW TRC samples to validate the presence or absence of the reported variation. Three Quantitative PCR (qPCR)-based CNV assays have been ordered to span the reported regions of variation (Labeled Assay 1 – Assay 3 in Figure 2.2). Identifying CNVs is crucial for interpreting association and expression analyses and will complete an initial analysis of *GPSM1* in human AD and craving-related behavior.
Figure 2.2. Reported CNVs in, and proximal to, *GPSM1*. Custom track visualized on the UCSC Genome Browser, http://genome.ucsc.edu (Kent, Sugnet et al. 2002). Headings on the left correspond to first author on papers reporting CNV, empirical data from the IASPSAD or CNV assay placement (see text). Base pair positions are based on the February 2009 GRch37/hg19 assembly.
CHAPTER THREE

Association of GRM3 and KIAA1324L with Alcohol Dependence in the Irish Affected Sib Pair Study of Alcohol Dependence

Introduction

Until recently, association studies for alcohol dependence (AD) have typically focused on genes involved in ethanol metabolism or neurotransmitter systems. As discussed in the introduction, evidence of association has been observed with alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH) genes as well as neurotransmitter candidates including the GABA receptor subunit gene GABRA2. While clearly representing some progress towards the identification of loci conferring risk for AD, both the lack of replication and modest strength of observed association with AD suggest that numerous genes contributing to the risk for AD remain to be identified.

The genome-wide association study (GWAS) has provided the opportunity to test for association unbiased by candidate gene selection. At the advent of the GWAS era, genotyping was prohibitively expensive (Meaburn, Butcher et al. 2006). As such, many of the first studies were pooled GWAS wherein genomic DNA is pooled together and run on a single genotyping array (Craig, Huentelman et al. 2005). Technical artifacts may arise in part because it cannot be ensured that each individual is equally represented in the pool. However, pooled GWAS studies can provide initial evidence of association signals if these can be redetected with genotyping at the individual level (Meaburn, Butcher et al. 2006, Kirov, Nikolov et al. 2006). A pooled GWAS of AD in the Collaborative Study on the Genetics of Alcoholism (COGA) sample reported association
with groups of genes involved in signaling, adhesion and gene regulation (Johnson, Drgon et al. 2006).

Previously, a pooled GWAS in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) case-control sample was conducted (unpublished data), identifying a cluster of 8 significantly associated single nucleotide polymorphisms (SNPs) on human chromosome 7q21.1-q21. The 8 SNPs map within and just upstream of a relatively uncharacterized locus, KIAA1324L (chr7:86,506,223-86,689,014), but are also in linkage disequilibrium (LD) with a metabotropic glutamate receptor and obvious candidate gene for AD, GRM3 (chr7:86,273,230-86,494,192). Finally, this region lies within an AD linkage region from COGA (Foroud, Edenberg et al. 2000) and an alcohol consumption quantitative trait loci (QTL) in a syntenic region from the rat (Carr, Foroud et al. 1998).

This chapter examines the cluster of LD-tagging SNPs at the individual genotyping level to confirm the results observed in the pooled study. SNPs in a 660kb region, encompassing both KIAA1324L and GRM3, were genotyped in the IASPSAD case-control sample and tested for association with AD and AD symptom count. Replication was sought in another ascertained sample and a population sample. Finally, KIAA1324L and GRM3 gene expression differences were examined in human chronic alcoholic and control post-mortem brain tissues.
Methods

Primary sample: Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD)

A detailed description of the sample is provided elsewhere (Prescott, Sullivan et al. 2006). Briefly, probands were ascertained in clinical treatment facilities and hospitals in the Republic of Ireland and Northern Ireland between 1998 and 2002. Probands were interviewed using an adapted version of the Semi-Structured Assessment of the Genetics of Alcoholism (SSAGA) and diagnosed with Alcohol Dependence (AD) using the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria. The sample is ethnically homogenous; a stipulation for inclusion in the study required each proband have all 4 grandparents born in Ireland, Northern Ireland, Scotland, Wales or England. Controls were recruited from the National Guard and army reserve in the Republic of Ireland and from volunteers donating at the Northern Ireland Blood Transfusion Service in Northern Ireland. Controls had no reported history of alcoholism. From the original linkage sample, 562 genetically independent AD cases and 569 controls were selected for genotyping analyses.

Replication sample: Collaborative Study on the Genetics of Alcoholism (COGA)

Probands for the COGA sample, details of which are described elsewhere (Dick, Plunkett et al. 2006), were recruited throughout the United States in 6 geographic locations from alcohol treatment programs. Large families were preferentially ascertained with a proband required to have 2 or more family members in the COGA sampling area and preferably sibships greater than 3. All individuals were given the SSAGA interview and AD diagnosed using DSM-IV criteria. From the larger COGA
sample, 847 genetically independent AD cases and 552 controls of European ancestry were used.

Replication sample: Molecular Genetics of Schizophrenia (MGS2) sample

Details of the sample recruitment are provided elsewhere (Sanders, Duan et al. 2008, Sanders, Levinson et al. 2010, Shi, Levinson et al. 2009). Briefly, a survey research company, Knowledge Networks, Inc. (KN), used random digit dialing to recruit 3364 adult, non-Hispanic, European American (EA) subjects from across the United States. Participants were given a questionnaire (nimhgenetics.org) with a modified Composite International Diagnostic Interview-Short Form (CIDI-SF), screening for lifetime diagnoses of major psychiatric disorders such as alcohol dependence (AD). The final dataset after initial screening measures, laboratory quality control, and post-GWAS quality control included 2357 European American controls (Shi, Levinson et al. 2009) which were used as a population sample in this study. All data (genotypes and phenotypes) are available by application to Database of genotypes and phenotypes (dbGaP, dpgap.ncbi.nlm.nih.gov, Study Accessions: phs000021.v2.p1 and phs000167.v1.p10 and DNA, LCLs and additional phenotypic data are available through the National Institute of Mental Health (NIMH) repository (nimhgenetics.org).

Pooled GWAS:

A pooled GWAS was conducted on 302 unrelated DSM-IV defined AD cases and 264 controls from the IASPSAD sample, a subset of the larger IASPSAD case-control sample. Case and control pools were created in triplicate and genotyped in triplicate on
the Affymetrix 500K arrays by the Translational Genomics Research Institute (TGEN) (http://www.tgen.org/). SNPs were ranked by silhouette scores (Lovmar, Ahlford et al. 2005), and the top 25,000 SNPs were selected and re-ranked by t-test scores and finally prioritized by genomic clustering.

**Single Nucleotide Polymorphisms and Genotyping: IASPSAD**

SNPs were genotyped using either multiplex or monoplex genotyping. Multiplex genotyping was performed on a GenomeLab SNPstream (Beckman Coulter, Fullerton, CA, USA) following manufacturer’s protocols, in panels of 12 SNPs matched for their extension type. SNP sequences were screened for repeats and homology with other genomic sequences prior to using the proprietary Beckman Coulter primer design program, Autoprimer, for constructing the multiplex panels. In instances where SNP sequences were repeat-rich or matching extension types were necessary for successful multiplex paneling, proxy SNPs with matching $r^2$ and minor allele frequency (MAF) criteria were substituted. SNPs failing in the first round of genotyping were re-paneled and SNPs failing twice or failing SNPstream QC parameters were dropped. Monoplex reactions were performed using Taqman Assays-on-Demand (Applied Biosystems, Foster City, CA, USA). One SNP, rs17616282, was genotyped using a custom Taqman Assay (sequence available upon request). Genotypes were called using the Analyst AD fluorescence detector (LJL Biosystems, Sunnyvale, CA, USA) and automated Excel template (van den Oord, Jiang et al. 2003). To minimize technical variability, all reaction steps were performed using the Eppendorf 5075 automated liquid handler. Stringent evaluation of initial data is important to avoid artificial effects of genotyping.
errors; therefore, all genotypes were independently assessed by two raters. Ambiguous calls were discussed and in cases of non-resolution, genotypes were dropped from the analyses.

Tagging SNPs (tSNPs) were selected using CEU individuals from HapMap (data Rel22, NCBI B36 assembly, dbSNP b126) and TAGGER (de Bakker, Yelensky et al. 2005) as implemented in Haplovew v3.2 (Barrett, Fry et al. 2005) using default criteria of MAF>0.2 and r²>0.8.

**Single Nucleotide Polymorphisms and Genotyping: replication samples**

Genotypes for analysis were extracted from the COGA and MGS2 GWAS datasets. As detailed elsewhere (Edenberg, Koller et al. 2010), COGA samples were genotyped on Illumina HumanHap1M Bead Chips (Illumina, San Diego, CA) by the Center for Inherited Disease Research (CIDR). After stringent quality control measures, 853,375 SNPs were available for analysis. MGS2 samples were genotyped at the Broad Institute on Affymetrix 6.0 arrays and, after thorough filtering, 696,788 SNPs were available for analysis (Shi, Levinson et al. 2009).

**Data analysis**

IASPSAD and COGA single marker association analyses and Hardy-Weinburg equilibrium (HWE) testing were performed using Plinkv1.07 (http://pngu.mgh.harvard.edu/purcell/plink) (Purcell, Neale et al. 2007). Logistic regression for AD diagnosis was performed on the IASPSAD and COGA case-control samples, with sex as a covariate. Linear regression with AD symptom count was
performed on the AD cases in the IASPSAD and COGA samples, with sex and age as covariates. SNPSpD (Nyholt 2004) was used to determine the number of independent tests for multiple testing corrections.

An AD factor score phenotype was constructed in the MGS2 control sample as detailed by Kendler et al (Kendler, Kalsi et al. 2011). Seven questions regarding symptoms of AD were extracted from the CIDI-SF questionnaire. These included craving, dangerous use, tolerance, loss of control, great period of time spent obtaining alcohol or recovering from its effects, activities given up and use despite knowledge of harm. Individuals who reported they never drank (n=498) were excluded from the analysis. Individuals who reported the most they drank was 1-3 drinks/day were given a score of 0 in subsequent analyses. All other individuals were given factor scores derived from the single factor solution reported in Kendler et al. Linear regression with the AD symptom factor score was performed on the MGS2 control samples in Plinkv1.07, with sex and age as covariates. SNPSpD was used to determine the number of independent tests for multiple testing correction.

Fisher’s combination test was used to test for replication (Peng, Luo et al. 2010).

Gene Expression Study: postmortem tissue sample

Superior frontal cortex (CTX), central nucleus of amygdala (CNA), and basolateral nucleus of amygdala (BLA) tissues from human chronic alcoholics and controls from the New South Wales Tissue Resource Centre (NSW TRC) located at the University of Sydney were used in this study. As described in Ponomarev, 2012, DSM-IV diagnosis of cases was based on clinical review. A GWAS was performed on 41
chronic alcoholics and 41 controls from the NSW TRC using Affymetrix 6.0 arrays (Riley et al, unpublished results). Expression data was available for 17 chronic alcoholics and 15 controls from the NSW TRC from a genomewide expression study using Illumina (Illumina, San Diego, CA) HumanHT-12 whole genome expression bead chips (Ponomarev, Wang et al. 2012). The samples from these studies were not congruent, so the overlap between the datasets, 12 chronic alcoholics and 8 controls, was chosen for analyses.

Gene expression study: statistical methods

Sex, age, PMI (post-mortem interval), liver pathology, neuropathology, toxicology, brain pH, and smoking demographics were provided by the NSW TRC. As 1 out of the 20 samples was female, sex was not included as a covariate. RNA integrity number (RIN) values were provided for each tissue, however a Fisher’s exact test was non-significant for differences between cases and controls in the CTX, CNA, and BLA (p=0.586, 1, and 0.906, respectively). RIN was therefore not included as a covariate. Liver pathology, neuropathology, and toxicology were re-coded numerically for subsequent analyses. Principal components analysis (PCA) was run in SASv9.3 using age, PMI, liver pathology, neuropathology, toxicology, brain pH, and smoking. The first 5 components explained 90% of the variance in the dataset (see Chapter Two, Figure 2.1). Component scores for the first 5 components were used as covariates in the analyses.

Four genes lying within the tagging region in the IASPSAD sample were chosen for inclusion in the study: GRM3 (Illumina identifiers ILMN_1679532 and
ILMN_2078975), *KIAA1324L* (Illumina identifier ILMN_1652371), *DMTF1* (Illumina identifiers ILMN_1750075 and ILMN_2119486), and *c7orf23* (Illumina identifier ILMN_1751143). Differences in gene expression between cases and controls were analyzed using the PROC GLM – MANOVA test in SASv9.3 and raw p-values corrected for the 6 probes tested. Probes with p-values surviving multiple testing correction were analyzed using ANOVA in each brain region using the PROC GLM test in SASv9.3. Linear regression tests for allele specific expression differences were performed in Plink v1.07 with the aforementioned component scores used as covariates.

**Results**

As described previously, the pooled GWAS of AD in the IASPSAD case-control sample identified a cluster of 8 significantly associated SNPs on human chromosome 7q21.1-q21.12 (unpublished data). This region was previously implicated in a COGA AD linkage scan (Foroud, 2000) and in a rat alcohol consumption QTL (Carr, Foroud et al. 1998). Since pooled GWAS datasets are susceptible to technical artifacts, the 8 SNPs were individually genotyped in the IASPSAD sample subset sent for pooled genotyping. The results were comparable between the pooled GWAS and individual genotyping (Table 3.1). All 8 SNPs remained nominally associated with AD.
Table 3.1. Validation of pooled GWAS results by individual genotyping in the IASPSAD.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Pooled GWAS Cs freq</th>
<th>Cn Freq</th>
<th>Individual Genotyping Cs freq</th>
<th>Cn Freq</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1635037</td>
<td>0.875</td>
<td>0.758</td>
<td>0.720</td>
<td>0.664</td>
<td>0.0055</td>
</tr>
<tr>
<td>rs1767741</td>
<td>0.744</td>
<td>0.620</td>
<td>0.720</td>
<td>0.666</td>
<td>0.0073</td>
</tr>
<tr>
<td>rs1767735</td>
<td>0.840</td>
<td>0.723</td>
<td>0.719</td>
<td>0.664</td>
<td>0.0065</td>
</tr>
<tr>
<td>rs767435</td>
<td>0.818</td>
<td>0.662</td>
<td>0.720</td>
<td>0.670</td>
<td>0.0135</td>
</tr>
<tr>
<td>rs767434</td>
<td>0.733</td>
<td>0.526</td>
<td>0.722</td>
<td>0.666</td>
<td>0.0059</td>
</tr>
<tr>
<td>rs1635010</td>
<td>0.899</td>
<td>0.779</td>
<td>0.722</td>
<td>0.668</td>
<td>0.0078</td>
</tr>
<tr>
<td>rs2373338</td>
<td>0.780</td>
<td>0.607</td>
<td>0.719</td>
<td>0.666</td>
<td>0.0086</td>
</tr>
<tr>
<td>rs1859122</td>
<td>0.761</td>
<td>0.624</td>
<td>0.780</td>
<td>0.723</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

All 8 SNPs remained nominally significant (p≤0.05) after individual genotyping in a subset of the IASPSAD sample. SNPs rs1635037-rs2373338 fall within a single LD block and are redundant.

The SNP cluster identified in the pooled GWAS lies within and upstream of *KIAA1324L*, and is in LD with a metabotropic glutamate receptor, *GRM3*. Therefore, a ~660kb region (chr7: 86039131-866992499) surrounding the initial cluster of associated SNPs, and encompassing a large block of LD, was analyzed for tagging SNPs (tSNPs). Thirty-two tSNPs were identified, 2 of which were already genotyped in the validation of the pooled results. Two SNPs were unavailable for both multiplex and monoplex genotyping. Five SNPs genotyped poorly and were dropped. Two SNPs showed deviation from HWE (p < 0.001) in control samples and were excluded. When combined with the 2 previously genotyped tSNPs, a total of 23 out of 30 available tSNPs (77%) genotyped successfully. Four additional SNPs were chosen for genotyping in the IASPSAD for consistency with the higher density GWAS data from the COGA and MGS replication samples, one of which genotyped poorly and was dropped.
Of the 26 successfully genotyped SNPs in the IASPSAD sample (23 tSNPs plus 3 additional SNPs), 1 SNP was removed for excessive missingness (0.13). Ninety-eight samples were removed for having ≥ 8 missing genotypes (≥32%). After data cleaning, a total of 25 tSNPs were analyzed for tests of association on 537 cases and 496 controls.

Table 3.2 shows the IASPSAD case-control sample association results for AD. A number of SNPs in the region were nominally associated with AD. SNPSpD (Nyholt 2004) was used to ascertain the number of independent SNPs for multiple testing correction. After correcting for 15 SNPs, rs802467 and rs1859122 remained significantly associated with AD (p<0.0034) and rs1635037 was borderline significant. Rs802467 and rs1635037 are located in introns of GRM3 and KIAA1324L, respectively. Rs802467 is in LD with SNPs in GRM3, only, and similarly rs1635037 is only in LD with SNPs in KIAA1324L. Rs1859122 maps ~44.5 kb upstream of KIAA1324L and is in weak LD (\(r^2<0.3\)) with SNPs in both KIAA1324L and GRM3.

Based on the association results in our sample, replication was attempted in an independent sample. Table 3.2 summarizes the association results for AD in the COGA sample. There was no evidence of association in COGA with SNPs previously implicated in the IASPSAD sample. However, one SNP within KIAA1324L and one SNP in DMTF1, distal to KIAA1324L, were nominally associated with AD in COGA. The most significant SNP, rs17609037, is only in LD with SNPs within KIAA1324L. SNPSpD identified 15 independent SNPs and after multiple testing correction, no SNPs remained significant (p<0.0034). Analysis of AD symptom count in the IASPSAD and COGA samples provided little additional association evidence (Table 3.3). Association results
Table 3.2. Case-control association results for AD in the IASPSAD and COGA.

<table>
<thead>
<tr>
<th>Gene</th>
<th>IASPSAD</th>
<th>COGA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNP</td>
<td>Alleles</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>A/G</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>A/G</td>
</tr>
<tr>
<td></td>
<td>A/G</td>
<td>86146536</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>86166546</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>86168009</td>
</tr>
<tr>
<td></td>
<td>C/T</td>
<td>86173748</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>86189153</td>
</tr>
<tr>
<td></td>
<td>T/C</td>
<td>86233862</td>
</tr>
<tr>
<td></td>
<td>C/T</td>
<td>86241147</td>
</tr>
<tr>
<td></td>
<td>C/T</td>
<td>86258176</td>
</tr>
<tr>
<td></td>
<td>C/T</td>
<td>86258313</td>
</tr>
<tr>
<td></td>
<td>A/T</td>
<td>86310267</td>
</tr>
<tr>
<td></td>
<td>A/G</td>
<td>86321515</td>
</tr>
<tr>
<td></td>
<td>A/G</td>
<td>86406457</td>
</tr>
<tr>
<td></td>
<td>A/G</td>
<td>86444443</td>
</tr>
<tr>
<td></td>
<td>G/A</td>
<td>86516142</td>
</tr>
<tr>
<td></td>
<td>G/A</td>
<td>86518964</td>
</tr>
<tr>
<td></td>
<td>C/T</td>
<td>86524186</td>
</tr>
<tr>
<td></td>
<td>G/A</td>
<td>86547029</td>
</tr>
<tr>
<td></td>
<td>C/T</td>
<td>86571409</td>
</tr>
<tr>
<td></td>
<td>C/T</td>
<td>86575011</td>
</tr>
<tr>
<td></td>
<td>G/A</td>
<td>86629940</td>
</tr>
<tr>
<td></td>
<td>A/G</td>
<td>86668126</td>
</tr>
</tbody>
</table>

Nominal p-values shown. An odds ratio (OR) > 1 indicates the minor allele (second in the "alleles" column) increases the odds of having AD. If the identical SNP genotyped in the IASPSAD sample was not available, proxy SNPs were analyzed (if available) and are noted in italics along with r² values. *SNP was significant after multiple testing correction.
Table 3.3. Association results for AD symptom count in the IASPSAD, COGA and MGS2 samples.

<table>
<thead>
<tr>
<th>Gene</th>
<th>IASPSAD</th>
<th>COGA</th>
<th>MGS2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SNP</td>
<td>Alleles</td>
<td>Position</td>
</tr>
<tr>
<td>rs1268991</td>
<td>T/C</td>
<td>-0.159</td>
<td>0.00999</td>
</tr>
<tr>
<td>rs274621</td>
<td>A/G</td>
<td>0.056</td>
<td>0.3548</td>
</tr>
<tr>
<td>rs13581638</td>
<td>A/G</td>
<td>-0.130</td>
<td>0.1340</td>
</tr>
<tr>
<td>rs1405875</td>
<td>T/C</td>
<td>-0.037</td>
<td>0.5246</td>
</tr>
<tr>
<td>rs2299218</td>
<td>C/T</td>
<td>-0.094</td>
<td>0.2190</td>
</tr>
<tr>
<td>rs10239714</td>
<td>T/C</td>
<td>-0.079</td>
<td>0.1998</td>
</tr>
<tr>
<td>rs6943659</td>
<td>T/C</td>
<td>-0.108</td>
<td>0.0723</td>
</tr>
<tr>
<td>rs6961677</td>
<td>C/T</td>
<td>-0.102</td>
<td>0.0968</td>
</tr>
<tr>
<td>rs10277739</td>
<td>C/T</td>
<td>-0.115</td>
<td>0.0717</td>
</tr>
<tr>
<td>rs1023440</td>
<td>T/C</td>
<td>-0.079</td>
<td>0.2610</td>
</tr>
<tr>
<td>rs7798115</td>
<td>A/T</td>
<td>-0.034</td>
<td>0.6723</td>
</tr>
<tr>
<td>rs7781414</td>
<td>A/G</td>
<td>-0.053</td>
<td>0.4126</td>
</tr>
<tr>
<td>KIAA1324L</td>
<td>A/G</td>
<td>0.043</td>
<td>0.4834</td>
</tr>
<tr>
<td>rs1760937</td>
<td>T/C</td>
<td>0.293</td>
<td>0.1111</td>
</tr>
<tr>
<td>rs1635037</td>
<td>C/T</td>
<td>-0.070</td>
<td>0.2765</td>
</tr>
<tr>
<td>KIAA1324L</td>
<td>A/G</td>
<td>0.044</td>
<td>0.4616</td>
</tr>
<tr>
<td>rs1176811</td>
<td>T/C</td>
<td>0.086</td>
<td>0.1304</td>
</tr>
<tr>
<td>rs1176811</td>
<td>T/C</td>
<td>0.085</td>
<td>0.2252</td>
</tr>
<tr>
<td>rs3859122</td>
<td>C/T</td>
<td>0.035</td>
<td>0.5787</td>
</tr>
<tr>
<td>rs1898611</td>
<td>C/A</td>
<td>0.062</td>
<td>0.4530</td>
</tr>
<tr>
<td>DMTF1</td>
<td>A/G</td>
<td>0.072</td>
<td>0.3548</td>
</tr>
<tr>
<td>rs17762292</td>
<td>A/G</td>
<td>0.016</td>
<td>0.7891</td>
</tr>
</tbody>
</table>

Nominal p-values shown. A positive regression coefficient (Beta) indicates the minor allele (second in the "alleles" column) increases the phenotype. If the identical SNP genotyped in the IASPSAD sample was not available, proxy SNPs were analyzed (if available) and are noted in italics along with $r^2$ values.
for an AD symptom factor score phenotype (see Methods) from a third independent sample, the MGS2 controls, are listed in the Table 3.3. There was no SNP-to-SNP replication between any of the samples. Two SNPs within GRM3 and one SNP within KIAA1324L were nominally associated in the MGS2 controls. The most significantly associated SNP, rs1999945, lies within KIAA1324L but is in moderate LD with SNPs within GRM3 ($r^2 \sim 0.64$). After correcting for 13 independent SNPs, no SNPs remained significantly associated ($p<0.0039$). Fisher's combination test was used to test for replication in the COGA and MGS2 datasets. The $p$-value of 0.115 (chi-square 7.46, df=4) was suggestive for replication, though not significant.

In an effort to clarify the primary risk locus in the associated region, expression of four genes in human postmortem brain tissues from the New South Wales Tissue Resource Center was examined. Expression was analyzed between chronic alcoholics ($n=12$) and controls ($n=8$) in three brain regions with implications for reward processing: the basolateral nucleus of amygdala (BLA), central nucleus of amygdala (CNA), and superior frontal cortex (CTX). Multivariate analysis of variance (MANOVA) results showed no expression differences for KIAA1324L, DMTF1 and c7orf23 (Table 3.4A). One probe for GRM3 showed significant expression differences between cases and controls after multiple testing correction. Separate analysis of variance (ANOVA) within each brain region showed significantly higher GRM3 expression in controls compared to cases (Table 3.4B and Figure 3.1). No allele-specific expression differences for SNPs surviving multiple testing correction were found for any of the four genes in the associated LD block (Table 3.5).
Table 3.4. Gene expression differences between chronic alcoholics and controls.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Illumina Identifier</th>
<th>p-value (corrected)</th>
<th>B. Tissue</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRM3</td>
<td>ILMN_1679532</td>
<td>0.2154</td>
<td>BLA</td>
<td>0.002</td>
</tr>
<tr>
<td>GRM3</td>
<td>ILMN_2078975</td>
<td>0.0027 (0.019)</td>
<td>CNA</td>
<td>0.031</td>
</tr>
<tr>
<td>KIAA1324L</td>
<td>ILMN_1652371</td>
<td>0.3419</td>
<td>CTX</td>
<td>0.003</td>
</tr>
<tr>
<td>DMTF1</td>
<td>ILMN_1750075</td>
<td>0.0430 (0.301)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMTF1</td>
<td>ILMN_2119486</td>
<td>0.3761</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c7orf23</td>
<td>ILMN_1751143</td>
<td>0.1957</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(A). MANOVA results for 4 genes of interest. Significant MANOVA p-values were corrected for the six tested probes. (B). GRM3 (ILMN_2078975) expression was analyzed in the basolateral nucleus of amygdala (BLA), central nucleus of amygdala (CNA), and superior frontal cortex (CTX). Raw ANOVA p-values are shown.

Table 3.5 Association results for allele-specific expression differences.

<table>
<thead>
<tr>
<th>Gene (Illumina probe)</th>
<th>BLA</th>
<th>CNA</th>
<th>CTX</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs802467</td>
<td>rs1635037</td>
<td>rs1859122</td>
<td></td>
</tr>
<tr>
<td>rs802467</td>
<td>rs1635037</td>
<td>rs1859122</td>
<td>rs802467</td>
</tr>
<tr>
<td>rs1635037</td>
<td>rs1859122</td>
<td></td>
<td>rs802467</td>
</tr>
<tr>
<td>rs1859122</td>
<td></td>
<td>rs802467</td>
<td>rs1635037</td>
</tr>
<tr>
<td>rs802467</td>
<td>rs1635037</td>
<td>rs1859122</td>
<td>rs802467</td>
</tr>
<tr>
<td>rs1635037</td>
<td>rs1859122</td>
<td></td>
<td>rs802467</td>
</tr>
<tr>
<td>rs1859122</td>
<td></td>
<td>rs802467</td>
<td>rs1635037</td>
</tr>
</tbody>
</table>

Association results for allele-specific expression differences in the NSW TRC samples. Bonferroni corrected p-values shown for each probe with 3 SNPs in 3 tested brain regions: the basolateral nucleus of amygdala (BLA), central nucleus of amygdala (CNA), and superior frontal cortex (CTX). P-values were corrected, within brain region, for 18 tests (3 SNPs x 6 probes). SNPs were chosen for analysis based on significant association results in the IASPSAD case-control sample. Rs802467 and rs1859122 were significantly associated with AD after multiple testing correction and rs1635037 was borderline significant.
Figure 3.1. Significant *GRM3* gene expression differences in human chronic alcoholic and control brain tissues. Box plots show expression differences in controls (1) and cases (2) in three brain regions: basolateral nucleus of amygdala (A), central nucleus of amygdala (B), and superior frontal cortex (C).
Discussion

On 7q21.1-q21.12, a region previously implicated in AD in humans and ethanol consumption in rats, three independent samples were analyzed for association with AD/AD symptom count. In the clinically ascertained IASPSAD sample, 2 SNPs were associated with AD after multiple testing correction. The minor allele of rs802467, intronic to GRM3 and only in LD with GRM3 SNPs, was associated with increased risk for AD. Replication was suggestive, though not significant, in another ascertained sample, COGA, and a population sample, the MGS2 controls. However, within the COGA dataset, the most significant SNP (rs17609037) was located in an intron of KIAA1324L, and only in LD with SNPs in that locus. Similarly, in the MGS2 controls, the most significant SNP (rs1999945) was located within an intron of KIAA1324L but was in moderate LD with SNPs in GRM3 as well. While positive association results in multiple samples support the existence of a gene or genes influencing AD risk within this genomic region, the lack of direct replication remains a limitation. Associated SNPs in one sample show no association in any other sample. This could be influenced by subtle differences in patterns of LD between the samples (Figures 3.2-3.4).
Figure 3.2. LD in the associated region of chromosome 7 in IASPSAD controls. Haploview shading based on $r^2$ values.

Figure 3.3. LD in the associated region of chromosome 7 in COGA controls of European descent. Haploview shading based on $r^2$ values.
To further investigate which gene in the region may contribute to risk for AD, the expression of the four genes in the associated LD block in chronic human alcoholic and control postmortem brain samples from the New South Wales Tissue Resource Centre was analyzed. A probe tagging the 3'UTR of GRM3 showed significant gene expression differences between cases and controls in the BLA, CNA and CTX. Expression was significantly higher in the control samples compared to the chronic alcoholic samples. No other genes in the region showed significant expression differences, though the small sample size (n=12 cases and n=8 controls) may have limited power to detect more subtle differences.
**GRM3** encodes a Group II metabotropic glutamate receptor (mGlu3); a G-protein coupled receptor that functions to inhibit adenylate cyclase and decrease levels of cyclic adenosine monophosphate (cAMP) (Lyon, Kew et al. 2008). Group II receptors (mGlu2 and mGlu3) also negatively regulate glutamate release (Gass, Olive 2008). Agonists of the Group II receptors result in lower levels of glutamate in the synapse (Cannady, Grondin et al. 2011). A GWAS on comorbid alcohol and nicotine dependent Australian and Dutch samples yielded one SNP within **GRM3**, rs218916, in the top 30 most significant SNPs (Lind, Macgregor et al. 2010). Although there has been substantial research on **GRM3** in other psychiatric and related phenotypes, there is little additional evidence specifically linking **GRM3** to AD or alcohol-related phenotypes in humans. However, a recent study by Xia and colleagues implicated **GRM3** in functional abnormalities seen in the prefrontal cortex of AD individuals (Xia, Ma et al. 2012).

Model organism research in rodents has yielded additional insight on the relationship between **GRM3** expression and alcohol related phenotypes. **Grm3** knockout mice showed no differences in response to ethanol (Corti, Andreoli et al. 2004), however **Grm3** is significantly ethanol responsive in the prefrontal cortex of mouse BXD Recombinant Inbred lines and is a hub in ethanol-responsive gene network analyses (Wolen, Phillips et al. 2012). This suggests that **Grm3** expression changes in response to ethanol exposure and may have a central role in ethanol responsive pathways in mice. In rats, activation of Group II receptors results in decreased ethanol seeking (Rodd, McKinzie et al. 2006), stress-induced (Sidhpura, Weiss et al. 2010) and cue-induced (Zhao, Dayas et al. 2006) relapse behavior and drinking (Sidhpura,
Weiss et al. 2010, Backstrom, Hyytia 2005). Taken together, results from rodents implicate Grm3 in ethanol response and ethanol-related behaviors.

However, the association seen with KIAA1324L, and whether or not it represents an independent signal, cannot be excluded. The KIAA1324L gene encodes a one-pass transmembrane protein that is evolutionarily conserved and widely expressed (Araki, Kusakabe et al. 2011). The *Xenopus laevis* ortholog of KIAA1324L, xEIG121L, has been implicated in bone morphogenetic protein (BMP) signaling (Araki, Kusakabe et al. 2011) and additional studies suggest alcohol decreases BMP-mediated Smad signaling in the liver of chronically exposed mice (Gerjevic, Liu et al. 2012). Though tenuous, the literature suggests that KIAA1324L may have a role in ethanol-related signaling.

This chapter investigated a region on human chromosome 7 associated with AD in the IASPSAD sample. Of the four genes located in the region, association and expression studies most strongly implicate GRM3. Evidence from the model organism literature implicates activation of Group II metabotropic glutamate receptors in “protective” alcohol-related behaviors and Grm3, specifically, in ethanol responsiveness. Though the signal is somewhat diffuse, KIAA1324L cannot be excluded as a potential risk locus for AD. KIAA1324L’s relationship with alcohol-related behaviors is not well documented, but is known to be involved in BMP signaling, which in turn may be altered due to alcohol exposure. Collectively, prior evidence and results from this chapter suggest that further research is warranted to investigate variation in and expression of GRM3 in human AD and related phenotypes.
Future Directions

Two final projects will be completed in an effort to clarify whether \textit{GRM3}, \textit{KIAA1324L} or both loci are risk factors for AD. First, an expression quantitative trait loci (eQTL) study is currently underway in the aforementioned NSW TRC samples using expression data from Ponomarev et al and GWAS data from Riley et al (see Methods). Initial results show significant evidence for two cis-eQTLs in \textit{GRM3}. Secondly, the recently released data from the Encyclopedia of DNA Elements (ENCODE) Consortium (ENCODE Project Consortium, Dunham et al. 2012) will be interrogated to see if any if the associated SNPs reported in this chapter are located in regulatory regions for \textit{GRM3} or \textit{KIAA1324L}. Preliminary review suggests that rs1859122 (significant after multiple testing correction in the IASPSAD sample) is less than 150bp away from an enhancer of \textit{KIAA1324L} (Figure 3.5). Completion of these two projects will strengthen the large body of evidence presented in this chapter linking this region of chromosome 7 to AD.
Figure 3.5. ENCODE graphic for the associated region on chromosome 7. All 9 SNPs nominally associated with AD in the IASPSAD study are represented as crosses. Rs1859122 (located on the far right of the graphic) is 5' to KIAA1324L and is located near an enhancer of that locus. (Figure courtesy of Dr. Aaron Wolen.)
Introduction

As discussed previously, alcohol dependence (AD) is a dichotomous clinical diagnosis. Researchers have had modest success finding genes associated with AD, but analyzing alcohol-related phenotypes (ARPs) is an attractive alternative. ARPs are specific factors though to contribute to the overall risk for AD. While AD is likely due in part to the action of numerous genes, ARPs, like physiological measurements, may be closer to the underlying genetic risk factors, may be localized to specific regions of the genome and are more likely to comprise smaller gene networks. Such a reduction in genetic complexity results in increased statistical power for the proposed association studies. Additionally, associated genes can be analyzed in the light of a relatively simpler biological component of a much more complex clinical phenotype.

Level of response to alcohol (LR) is a powerful indicator of future alcohol use problems, alcohol use disorders and alcohol consumption. Individuals with low LR feel less effect from each drink and are at an increased risk for alcohol use problems, typically drinking more heavily than individuals with high LR (Schuckit, Smith et al. 2012, Schuckit, Smith et al. 2007, Schuckit, Smith et al. 2011, Schuckit, Smith et al. 2009, Schuckit, Smith 2001, Trim, Schuckit et al. 2009, Daeppen, Landry et al. 2000, Schuckit, Smith et al. 2009, Schuckit, Smith et al. 2008). This correlation has been documented across sexes (e.g., men (Schuckit, Smith et al. 2004) and women
(Schuckit, Smith et al. 2003), ages (e.g., adolescents (Schuckit, Smith et al. 2005)), and races (e.g., African Americans (Pedersen, McCarthy 2009).

A direct alcohol challenge can be given to experimental subjects to measure their level of response to alcohol. However, this test is expensive and time consuming. Schuckit et al developed a (SRE) instrument that provides information on the same physiological measurements, but is a paper survey and takes less than five minutes to complete (Schuckit, Tipp et al. 1997, Schuckit, Smith et al. 1997). When the SRE was given to ninety-eight men who also completed the alcohol challenge test, the correlation between the results was 0.82 (p<0.001) (Schuckit, Tipp et al. 1997). The SRE asks “how many drinks did it take for you to begin to feel different, feel a bit dizzy or begin to slur your speech, begin stumbling or walking in an uncoordinated manner, or pass out or fall asleep when you did not want to?” during the first five times the subject drank. An initial sensitivity (ISENS) score is computed based on the number of drinks reported divided by the number of effects endorsed.

LR is heritable with estimates from twin studies ranging from 40-60% (Heath, Madden et al. 1999, Viken, Rose et al. 2003, Martin, Oakeshott et al. 1985). Among first degree relatives, the LR correlation is reported to be between 0.14 and 0.22 (Schuckit, Smith et al. 2005). Furthermore, children of alcoholics, when compared to other family members, have lower LR (de Wit, McCracken 1990, Eng, Schuckit et al. 2005). Given the heritability of the trait and the predictive ability of low LR for future alcohol use problems, there has been substantial interest in identifying its genetic risk factors.
Multiple linkage studies have examined LR. In a population of American Indians, strong linkage was seen on chromosomes 6 and 9 (Ehlers, Gizer et al. 2010). The Collaborative Study on the Genetics of Alcoholism (COGA) found linkage with \textit{KCNMA1}, \textit{HTR7} and \textit{SLC18A2} (Schuckit, Wilhelmsen et al. 2005) and on chromosomes 11, 13, 20 and 21 (Schuckit, Edenberg et al. 2001). In a U.S. sample from San Diego, linkage was found on chromosomes 10, 11 and 22 (Wilhelmsen, Schuckit et al. 2003) and with the \textit{CYP2E1} locus (Webb, Lind et al. 2011). In the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD), linkage peaks were identified on chromosomes 1 and 11 (Kuo, Neale et al. 2006).

The IASPSAD ISENS linkage peak on chromosome 1 is particularly intriguing. The ISENS peak at D1S2726 (LOD=1.89, \textit{p}-value=0.002) is paired with a proximal alcohol tolerance peak at D1S219 (LOD 1.78, \textit{p}-value=0.002) (Kuo, Neale et al. 2006). The 41Mb region between the peaks (chr1: 69,841,001-111,184,618) has been implicated in three other independent, alcohol-related linkage studies. The first sample, comprised of 18 UK families, showed linkage to 1p22.1-11.2 (peak at D1S1588, LOD 1.8) using the Research Diagnostic Criteria for Alcoholism (RDCA) as the phenotype (Guerrini, Cook et al. 2005). Second, 87 trios (proband and parents) obtained in Connecticut peaked at D1S406 (\textit{p}=0.005) using a TDT test with AD as the phenotype (Lappalainen, Kranzler et al. 2004). Third, COGA has repeatedly found evidence for linkage to chromosome 1, a brief summary of which is presented below.

Data from the initial COGA sample gave evidence of linkage for AD to D1S532 and D1S1588 (\textit{p}=0.0003) (peaks at D1S1588, LOD 2.93 and D1S224, LOD 1.65) (Reich, Edenberg et al. 1998). After genotyping an additional 105 families, and
obtaining a replication sample of 157 families, the COGA data was reanalyzed. The new linkage peaks for AD were located at D1S1588 (LOD 2.5) and D1S224/D1S1665 (LOD 1.7) (Foroud, Edenberg et al. 2000). The aforementioned replication sample peaked at D1S224 (LOD 1.6) using ICD-10 AD criteria. When both samples were analyzed together (original COGA data with 105 new families and the replication sample), there was a linkage peak at D1S2614/D1S1588 (LOD 2.6). The COGA data was further analyzed with alternative phenotypes for AD. SRE scores for ISENS in 745 sib pairs showed a linkage peak at D1S224 (LOD 1.9) (Schuckit, Wilhelmsen et al. 2005, Schuckit, Edenberg et al. 2001). AD and major depression (MD) analyzed together peaked at D1S1648/S1S1588 (LOD 5.12) while an “AD or MD” phenotype peaked at D1S224 (LOD 4.66) (Nurnberger, Foroud et al. 2001). Lastly, when the data were analyzed with a sex-and age-adjusted multiple threshold liability (MTL) model, two peaks were found: D1S532 (LOD 5.17) and D1S1665 (LOD 2.61). The MTL model included persons with AD and controls and used the differential severity of different diagnostic scales (Corbett, Saccone et al. 2005). In summary, four separate samples, encompassing thirteen different analyses, have found significant evidence of linkage to the same region on chromosome 1 for AD or ARPs. These results suggest the presence of a significant region for ISENS and ARPs surrounding the IASPSAD linkage peak on chromosome 1.

Human candidate gene studies have found association between LR and GABRG1 (Ray, Hutchison 2009), GABRA6 (Schuckit, Mazzanti et al. 1999), GAD1 (Kuo, Kalsi et al. 2009), SLC6A4 (Hinckers, Laucht et al. 2006) and the long allele of HTTLPR (Hu, Oroszi et al. 2005). Rs1051730 on chromosome 15 has been implicated.
in LR. It is located in a dense block of linkage disequilibrium (LD) but obvious candidate
genes in the region are CHRNA5, CHRNA3 and CHRNB4 (Joslyn, Brush et al. 2008). A recent genome-wide analysis found overrepresentation of neuronal signaling genes
associated with LR (Joslyn, Ravindranathan et al. 2010).

ISENS is a heritable, alcohol-related phenotype consistently correlated with risk for AD. Understanding the genetic risk factors for initial sensitivity to alcohol would be a powerful therapeutic tool. This chapter presents initial data from a genome-wide
association study (GWAS) on ISENS in the IASPSAD sample. Single nucleotide polymorphism (SNP) association analysis and gene-based association analysis are used to identify loci associated with risk for the phenotype.

Methods

Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD):

A detailed description of the sample is provided elsewhere (Prescott, Sullivan et al. 2006). Briefly, probands were ascertained in clinical treatment facilities and hospitals in the Republic of Ireland and Northern Ireland between 1998 and 2002. Probands were interviewed using an adapted version of the Semi-Structured Assessment of the Genetics of Alcoholism (SSAGA) and diagnosed with Alcohol Dependence (AD) using criteria from the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). The sample is ethnically homogenous; a stipulation for inclusion in the study required each proband have all 4 grandparents born in Ireland, Northern Ireland, Scotland, Wales or England.
Initial Sensitivity Phenotype

Initial Sensitivity (ISENS) was measured with the Self-Rating for the Effects of Alcohol (SRE) instrument developed by Schuckit et al (Schuckit, Tipp et al. 1997). ISENS is defined as the number of drinks needed for a person to feel the effects of ethanol during their first five times he or she drinks. The SRE queries four different effects and asks “how many drinks did it take for you to begin to feel different, feel a bit dizzy or begin to slur your speech, begin stumbling or walking in an uncoordinated manner, or pass out or fall asleep when you did not want to?” The phenotypic score for each individual was calculated as the number of total drinks reported divided by the total number of symptoms endorsed.

GWAS: pre-genotyping QC of IASPSAD cases

DNA from all probands and affected siblings from the IASPSAD was quantified using both spectrophotometry and Quant-iT Picogreen (Life Technologies, Carlsbad, CA) technology. 200ng of each DNA sample was run on a 2% agarose gel to assess degradation. Finally, all samples were sex-typed with sex chromosome-specific quantitative PCR (qPCR) using the Quantifiler Duo DNA Quantification Kit (Life Technologies, Carlsbad, CA). Only high molecular weight samples (bands > 800bp) with DNA concentrations ≥ 50ng/ul and with genotypic sex matching phenotypic sex were sent for genotyping.
815 samples from the IASPSAD, including eleven sets of parents to assess Mendelian errors, were sent to Vanderbilt University for genotyping on Affymetrix v6.0 arrays (Santa Clara, CA). 2000 Biobank controls from the Irish Blood Transfusion Service were also genotyped on Affymetrix v6.0 arrays at either the Affymetrix or Broad Institute core facilities. Arrays were called used the Affymetrix Birdseed v2 algorithm. Thirty-five samples were removed post-genotyping for CQC values <0.4, Nsp/Sty CQC values < 0 or discrepant genotypic sex calls. Arrays were then re-called using BEAGLECALL software (Browning, Yu 2009) in an effort to minimize site differences in genotyping. BEAGLECALL is an iterative process which starts with the fluorescent intensity calls and then incorporates haplotypic information from the sample with increasingly stringent calling filters. Subsequently, samples and SNPs underwent a thorough QC process. Samples were removed if the phenotypic and genotypic sex were discordant, they had excessive Mendelian errors, they were ancestry, heterozygosity or mean pi-hat outliers, their call rates were ≤ 98%, or if there was a DNA mix-up. SNPs were removed if Hardy-Weinburg equilibrium (HWE) p-values < 1x10^-6, the minor allele frequency (MAF) < 1%, there were Mendelian errors, they were duplicates, their call rate < 98% or if there was differential missingness between cases and controls. After QC, 676,736 SNPs were available for 710 probands and affected siblings and 1755 controls.
Imputation

Genotypes were imputed using IMPUTE2 (Howie, Donnelly et al. 2009) with the March 2012 1000 Genomes reference haplotype panel (1000 Genomes Project Consortium 2010). Probabilities were converted to dosages (based on number of minor alleles) using MACH2 (Scott, Mohlke et al. 2007). Post-imputation filtering removed SNPs with MAF < 1% and information content ≤0.3, leaving approximately 8.2 million SNPs.

Statistical Analyses

Phenotype distribution was visualized using PASW 18 (SPSS Inc). Association analyses for ISENS were performed within AD cases using ProbABEL (Aulchenko, Struchalin et al. 2010) in R 2.14.1 (R Core Team 2012). ProbABEL is unique in its ability to deal with related individuals, imputed genotypes and quantitative traits. To account for relatedness in the IASPSAD sample, ProbABEL calculates a kinship matrix. It then uses a two-step mixed model to perform linear regression. Sex and age at interview were included as covariates and residuals were used in the regression analysis. P-values were calculated using the Wald test. Based on the estimated 1.6 million independent tests in the 1000 Genomes dataset, genome-wide significance was set at $3.06 \times 10^{-8}$ (Li, Yeung et al. 2012). The false discovery rate (FDR) was calculated using the QVALUE package (Storey, Tibshirani 2003) in R (R Core Team 2012). Gene-based analyses were performed using the Versatile Gene-Based test for Genome-wide Association Studies (VEGAS) program. The program assigns SNPs to genes, accounts for LD through a series of Monte Carlo simulations and calculates an
empirical gene-based test statistic for each gene (Liu, McRae et al. 2010). The authors suggest a Bonferroni-corrected significance threshold of $p<2.83 \times 10^{-6}$ (17,787 genes tested).

**Results**

Figure 4.1 shows the distribution of ISENS values in the IASPSAD GWAS sample. ISENS scores range from 1 to 27.8 with a mean of 6.2 drinks to feel an effect. Though there is a wide distribution of values, overall the values are quite high, representative of the severely affected nature of the ascertained AD cases in the IASPSAD sample.

Figures 4.2 and 4.3 show the QQ plot and Manhattan plot, respectively, for ISENS association analyses. Nineteen SNPs were genome-wide significant (Table 4.1) and FDR $q$-values suggest a negligible false positive rate among them. The top SNP, rs143419076 ($p=2.7 \times 10^{-12}$), is intronic to GPLD1 on chromosome 6. Of the remaining eighteen SNPs, thirteen are intergenic. Rs184237132, rs75163186, rs189396701, rs11553355 and rs138694682 are intronic to ABCA9, AADACL4, STOX, CERKL/ITGA4 and KIAA0368, respectively. Gene-based analysis yielded no genes passing Bonferroni correction ($p<2.8 \times 10^{-6}$). Thirteen genes had $q$-values $\leq 0.11$ and all were located in the same region of chromosome 5 in one of two clusters (Table 4.2).
Figure 4.1. Distribution of ISENS values in the IASPSAD (n=680).
Figure 4.2. QQ plot for ISENS. The red line represents distribution under the null hypothesis of no association. $\lambda = 0.999$.

Figure 4.3. Manhattan plot for ISENS. The red line represents genome-wide significance at $p \leq 3.06 \times 10^{-8}$.
Table 4.1. SNPs associated with ISENS surviving multiple testing correction.

<table>
<thead>
<tr>
<th>Chr.</th>
<th>SNP</th>
<th>Position (bp)</th>
<th>MAF</th>
<th>Gene</th>
<th>p-Value</th>
<th>q-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>rs143419076</td>
<td>24489343</td>
<td>0.013</td>
<td>GPLD1</td>
<td>2.67E-12</td>
<td>2.2E-05</td>
</tr>
<tr>
<td>1</td>
<td>rs144574114</td>
<td>106314135</td>
<td>0.011</td>
<td></td>
<td>1.90E-11</td>
<td>8.0E-05</td>
</tr>
<tr>
<td>1</td>
<td>rs4526626</td>
<td>12744893</td>
<td>0.011</td>
<td></td>
<td>3.56E-11</td>
<td>0.0001</td>
</tr>
<tr>
<td>2</td>
<td>rs116150747</td>
<td>182319057</td>
<td>0.019</td>
<td></td>
<td>3.25E-10</td>
<td>0.0006</td>
</tr>
<tr>
<td>17</td>
<td>rs184237132</td>
<td>67041220</td>
<td>0.011</td>
<td>ABCA9</td>
<td>3.77E-10</td>
<td>0.0006</td>
</tr>
<tr>
<td>1</td>
<td>rs75163186</td>
<td>12712832</td>
<td>0.016</td>
<td>AADACL4</td>
<td>1.77E-09</td>
<td>0.0021</td>
</tr>
<tr>
<td>2</td>
<td>rs188552436</td>
<td>137266178</td>
<td>0.010</td>
<td></td>
<td>1.78E-09</td>
<td>0.0021</td>
</tr>
<tr>
<td>1</td>
<td>rs75497232</td>
<td>12761116</td>
<td>0.016</td>
<td></td>
<td>2.48E-09</td>
<td>0.0023</td>
</tr>
<tr>
<td>1</td>
<td>rs140269672</td>
<td>12749493</td>
<td>0.016</td>
<td></td>
<td>2.58E-09</td>
<td>0.0023</td>
</tr>
<tr>
<td>10</td>
<td>rs189396701</td>
<td>70592736</td>
<td>0.011</td>
<td>STOX1</td>
<td>4.49E-09</td>
<td>0.0037</td>
</tr>
<tr>
<td>1</td>
<td>rs61813510</td>
<td>156583538</td>
<td>0.025</td>
<td></td>
<td>6.64E-09</td>
<td>0.0049</td>
</tr>
<tr>
<td>17</td>
<td>rs138109588</td>
<td>35133526</td>
<td>0.012</td>
<td></td>
<td>1.34E-08</td>
<td>0.0092</td>
</tr>
<tr>
<td>1</td>
<td>rs145567095</td>
<td>90838991</td>
<td>0.011</td>
<td></td>
<td>1.59E-08</td>
<td>0.0100</td>
</tr>
<tr>
<td>2</td>
<td>rs11553355</td>
<td>182401587</td>
<td>0.023</td>
<td>CERKL/ITGA4</td>
<td>1.86E-08</td>
<td>0.0100</td>
</tr>
<tr>
<td>3</td>
<td>rs138099338</td>
<td>82488453</td>
<td>0.013</td>
<td></td>
<td>2.05E-08</td>
<td>0.0100</td>
</tr>
<tr>
<td>9</td>
<td>rs138694682</td>
<td>114240446</td>
<td>0.015</td>
<td>KIAA0368</td>
<td>2.17E-08</td>
<td>0.0100</td>
</tr>
<tr>
<td>1</td>
<td>rs149656664</td>
<td>106411940</td>
<td>0.013</td>
<td></td>
<td>2.25E-08</td>
<td>0.0100</td>
</tr>
<tr>
<td>16</td>
<td>rs75885089</td>
<td>51064968</td>
<td>0.035</td>
<td></td>
<td>2.33E-08</td>
<td>0.0100</td>
</tr>
<tr>
<td>2</td>
<td>rs181468506</td>
<td>128573560</td>
<td>0.031</td>
<td></td>
<td>2.43E-08</td>
<td>0.0100</td>
</tr>
</tbody>
</table>
Table 4.2. Results for ISENS gene-based association.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Gene</th>
<th>Start</th>
<th>Stop</th>
<th>No. SNPs</th>
<th>p-Value</th>
<th>TopSNP</th>
<th>SNP p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>HARS</td>
<td>140033673</td>
<td>140051155</td>
<td>90</td>
<td>4.80E-05</td>
<td>rs2531359</td>
<td>1.92E-06</td>
</tr>
<tr>
<td>5</td>
<td>HARS2</td>
<td>140051201</td>
<td>140059074</td>
<td>75</td>
<td>2.30E-05</td>
<td>rs2531348</td>
<td>1.75E-06</td>
</tr>
<tr>
<td>5</td>
<td>ZMAT2</td>
<td>140060215</td>
<td>140066423</td>
<td>70</td>
<td>2.10E-05</td>
<td>rs2531348</td>
<td>1.75E-06</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA1</td>
<td>140146059</td>
<td>140372113</td>
<td>227</td>
<td>3.20E-05</td>
<td>rs2531348</td>
<td>1.75E-06</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA2</td>
<td>140154627</td>
<td>140372113</td>
<td>215</td>
<td>5.40E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA3</td>
<td>140160966</td>
<td>140372113</td>
<td>214</td>
<td>5.50E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA4</td>
<td>140166855</td>
<td>140372113</td>
<td>210</td>
<td>6.20E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA5</td>
<td>140181544</td>
<td>140372113</td>
<td>206</td>
<td>5.00E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA6</td>
<td>140187833</td>
<td>140372113</td>
<td>202</td>
<td>6.50E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA7</td>
<td>140194152</td>
<td>140372113</td>
<td>197</td>
<td>6.00E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA8</td>
<td>140201090</td>
<td>140372113</td>
<td>185</td>
<td>6.50E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA9</td>
<td>140207540</td>
<td>140372113</td>
<td>179</td>
<td>7.60E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
<tr>
<td>5</td>
<td>PCDHA10</td>
<td>140215817</td>
<td>140372113</td>
<td>166</td>
<td>9.40E-05</td>
<td>rs991918</td>
<td>1.79E-05</td>
</tr>
</tbody>
</table>

All genes have q-value < 0.11.

Discussion

This chapter presents results from a GWAS of ISENS in the IASPSAD sample. There were a wide distribution of ISENS values in the sample, but most were quite high with a mean of 6.2 and a maximum of 28 drinks needed to feel an effect. The GWAS identified nineteen genome-wide significant SNPs. The top SNP, rs143419076 (p=2.7x10^{-12}), is intronic to GPLD1. GPLD1, Glycosylphosphatidylinositol specific phospholipase D1, is an enzyme which degrades glycosylphosphatidylinositol (GPI).
anchors, detaching proteins from the plasma membrane. It is predominately expressed in the liver (Schofield, Rademacher 2000). The actions of GPLD1 are involved in second messenger signaling, protein level homeostasis and protein activity (Nosjean, Briolay et al. 1997). Rs18834155, 5’ to the gene, and rs9467160, intronic to GPLD1, are associated with liver enzymes and liver disease ((Chambers, Zhang et al. 2011) and (Yuan, Waterworth et al. 2008), respectively). No reports linking the gene to initial sensitivity were found in the literature. However, an oligonucleotide array study in the Stanley Foundation Brain Bank sample found GPLD1 messenger RNA (mRNA) expression was increased in the middle temporal gyrus of alcoholics compared to controls (Sokolov, Jiang et al. 2003).

Five additional genome-wide significant SNPs in the GWAS implicate six additional loci in ISENS in our sample. AADACL4 and STOX1 have no known link, direct or indirect, to ISENS or AD. AADACL4, arylacetamide deacetylase-like, is uncharacterized in humans. STOX1, storkhead box 1, is a winged helix domain containing transcription factor and is a risk locus for pre-eclampsia (van Dijk, van Bezu et al. 2010). KIAA0368 and ABCA9 have some evidence of involvement in alcohol-related phenotypes. KIAA0368, proteasome-associated protein ECM29 homolog, associates with the 26S proteasome and is hypothesized to direct it to cellular locations needing rapid protein degradation (Gorbea, Goellner et al. 2004). In mice, Kiaa0368 is correlated with brain levels of the neurosteroid Deoxycorticosterone (DOC). Interestingly, DOC levels in mice increase after acute ethanol administration and decrease with ethanol dependence (Porcu, O’Buckley et al. 2011). ABCA9, ATP-binding cassette sub-family A member 9, encodes a protein which transports molecules
across cellular membranes and is expressed in the brain (Piehler, Kaminski et al. 2002). The exact function of ABCA9 is yet to be determined, but it is hypothesized to be involved in lipid transport (Piehler, Ozcurumez et al. 2012). In a recent report, Bettinger et al implicated changes in the lipid environment of Caenorhabditis elegans in the acute tolerance to ethanol (Bettinger, Leung et al. 2012).

The remaining genes implicated in the GWAS were CERKL and ITGA4. The genome-wide significant SNP rs11553355 is located in the 3-UTR of both genes. The function of CERKL, ceramide kinase-like, is still uncharacterized, however CERKL mutations are associated with cone-rod dystrophy (CRD) and recessive, nonsyndromic retinitis pigmentosa (RP26) (Mandal, Tran et al. 2012). ITGA4, integrin alpha 4, is a member of the integrin family. Integrins function as membrane spanning adhesion receptors. The protein encoded by ITGA4, the α4 chain, dimerizes with one of two beta chains. The α4β1 integrin is integral for hematopoiesis (Arroyo, Yang et al. 1999) and is associated with central nervous system (CNS) inflammation, migration of lymphocytes (Correia, Coutinho et al. 2009), autism (Correia, Coutinho et al. 2009) and poor prognosis in leukemia (Shanafelt, Geyer et al. 2008). Interestingly, ITGA4 is located in a significant quantitative trait locus (QTL) for acute ethanol sensitivity in BXD recombinant inbred mouse lines (Guo, Webb et al. 2009, Palmer, Lessov-Schlaggar et al. 2006), a phenotype directly comparable to the ISENS phenotype analyzed in our human sample.

In addition to the GWAS SNP results, gene-based association analyses yielded two clusters of genes on chromosome 5 associated with ISENS. One cluster is comprised of ZMAT2, HARS2, HARS and DND1, none of which have known relevance
to ISENS, AD or addiction. The second cluster is comprised of ten protocadherin isoform precursors, *PCDH1-10*. Protocadherins are integral membrane and adhesion proteins that function within the brain as cell connectors (Sano, Tanihara et al. 1993). In mice, multiple PCDH isoforms were reportedly associated with cocaine exposure and withdrawal: PCDH1, PCDH5 and PCDH10 expression increased during cocaine exposure and withdrawal while PCDH2 and PCDH6-9 expression increased during cocaine withdrawal, only (Eipper-Mains, Kiraly et al. 2012). There are no reports in the literature for involvement of these protocadherins in initial sensitivity to alcohol. However, PCDH9 is located within a QTL for chronic alcohol withdrawal in mice (Guo, Webb et al. 2009, Bergeson, Kyle Warren et al. 2003).

The thirteen remaining genome-wide significant SNPs are intergenic. Three of these SNPs (rs144574114, rs145567095 and rs149656664) are located within the ARP linkage region in the IASPSAD sample on chromosome 1. In fact, two of the three SNPs (rs144574114 and rs149656664) are located specifically within the initial sensitivity linkage peak itself (chr1: 101,685,305-111,184,618). Further research is necessary to examine whether the SNPs lie within DNA regulatory regions, microRNAs or are in LD with any known loci.

In summary, a GWAS of ISENS yielded multiple genome-wide significant SNPs. Six SNPs were located in known loci. *AADACL4, STOX1, GPLD1, KIAA0368, ABCA9* and *CERKL* represent new potential risk loci for ISENS. The final gene, *ITGA4*, has been implicated in acute sensitivity to ethanol in mice but this is the first known report of an association in humans. Gene-based analysis also implicated several loci in ISENS including a family of protocadherins that had been previously linked with addiction and...
alcohol withdrawal. The IASPSAD is a severely affected, ascertained sample and therefore these results need to be explored in a population sample to determine applicability. The associated SNPs from the GWAS had low MAF, ranging from 1-4%. Therefore, these results may need to be interpreted with caution. The IASPSAD sample size limits power to detect these smaller effect sizes and it is possible that a small number of genotypes could be artificially inflating the association results. However, in several instances the genome-wide significant SNPs were clustered in the same chromosomal region, indicating a true signal and arguing against artificial inflation. Replication in other human datasets would strengthen the association results and molecular characterization is necessary to untangle the role of each associated loci in the ISENS phenotype in our sample.

**Future Directions**

These results represent the initial GWAS analysis of the ISENS phenotype in the IASPSAD sample. There are multiple additional studies to be undertaken. A thorough analysis of SNPs in LD with the top SNPs is integral to analysis of the data. Top SNPs in LD with other associated SNPs are more indicative of a true association. Genes associated with ISENS in our sample will be checked for association with AD as well. As mentioned previously, replication will be attempted in COGA and other samples to strengthen the association results. Candidate genes from previous studies of ISENS will be examined in the data, possibly by weighting these results and re-analyzing the data. For significantly associated loci, messenger RNA (mRNA) differences will be assessed in postmortem brain samples from chronic alcoholics and controls. Also,
Association List Go AnnoTatOR (ALIGATOR) will be run in the dataset to identify pathways and Gene Ontology (GO) groups that are overrepresented (Holmans, Green et al. 2009).

Finally, a major strength of physiological phenotypes in human research is their application to model organisms. The results from this analysis have not been explored in depth in the model organism literature to see what existing data reveals. It may also be worthwhile to examine genes of interest in model organisms represented within the VCU Alcohol Research Center. Particularly in *Drosophila melanogaster* and *Caenorhabditis elegans*, expression of loci can be knocked down (or knocked out) and changes in acute tolerance and initial sensitivity reliably measured.
A Genome-wide Association Study of Maximum Alcohol Consumption in the Irish Affected Sib Pair Study of Alcohol Dependence

Introduction
Maximum alcohol consumption in a twenty-four period (MAX24) is a physiological phenotype related to Alcohol Dependence (AD) and is positively correlated with risk for alcohol use disorders (Dawson, Grant et al. 2005). Schuckit et al have found that the maximum number of drinks in a day was fifty percent higher among AD cases with withdrawal and tolerance (characterized as a more severe class) than AD cases endorsing neither (Schuckit, Tipp et al. 1995). Findings from several studies have shown that MAX24 is heritable. In a twin study of US Vietnam veterans, heritability was estimated at around thirty-five percent (Slutske, True et al. 1999). Saccone et al reported that in an unpublished adult twin study by Heath et al the heritability was estimated to be fifty percent (Saccone, Kwon et al. 2000). Furthermore, in the Collaborative Study of the Genetics of Alcoholism (COGA), the correlation between siblings was indicative of a genetic component to the MAX24 phenotype (~0.4 in same-sex sibling pairs) (Saccone, Kwon et al. 2000).

Previous genetic analyses of maximum alcohol consumption have implicated numerous chromosomal regions and candidate loci. Linkage analyses in the COGA sample found linkage to a region on chromosome 4 including the ADH gene cluster (Saccone, Kwon et al. 2000), a region on chromosome 7 encompassing TAS2R38 and
CHRM2 (Saccone, Saccone et al. 2005) and to chromosomes 10 and 13 (de Andrade, Olswold et al. 2005). In the Framingham Heart Study, linkage was found with markers on chromosomes 1,4,6,7,9,12,15,16,17 and 22. The top hit was on chromosome 9 near TRPM3 (Bergen, Yang et al. 2003). Linkage analysis in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) implicated chromosomes 12 and 18 (Kuo, Neale et al. 2006). Two candidate gene studies have examined association with alcohol metabolism genes and the MAX24 phenotype. In the Collaborative Genetic Study of Nicotine Dependence (COGEND), association was found with single nucleotide polymorphisms (SNPs) in ADH1B, ADH7, ALDH5A, ALDH1A2, and ALDH1A3 (Sherva, Rice et al. 2009). MacGregor et al found association with rs1229984 within ADH1B (Macgregor, Lind et al. 2009). A non-synonomous coding SNP within a bitter taste receptor gene, hTAS2R16, is also reported to be associated with AD and maximum alcohol consumption (Hinrichs, Wang et al. 2006). Finally, two genome-wide association studies (GWAS) of MAX24 found significant association with rs2140418, intronic to ANKS1A (Heath, Whitfield et al. 2011) and two SNPs within ASKRD7 (Chen, Xiong et al. 2012).

MAX24 is an alcohol-related phenotype shown to be associated with the severity and prevalence of alcohol use disorders. In light of these facts, and because of the documented heritability of the trait, we examined MAX24 in our sample. This chapter presents data from a GWAS on MAX24 in the IASPSAD sample. Single SNP association analysis and gene-based association analysis are used to identify loci associated with risk for the phenotype.
Methods

Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD)

A detailed description of the sample is provided elsewhere (Prescott, Sullivan et al. 2006). Briefly, probands were ascertained in clinical treatment facilities and hospitals in the Republic of Ireland and Northern Ireland between 1998 and 2002. Probands were interviewed using an adapted version of the Semi-Structured Assessment of the Genetics of Alcoholism (SSAGA) and diagnosed with Alcohol Dependence (AD) using DSM-IV criteria. The sample is ethnically homogenous; a stipulation for inclusion in the study required each proband have all 4 grandparents born in Ireland, Northern Ireland, Scotland, Wales or England.

Maximum Drinking Phenotype

Maximum alcohol consumption (MAX24), the largest number of drinks an individual reported consuming in 24 hours, was extracted from the SSAGA. Because units in Ireland differ from those in the United States, participants were asked to describe the type of beverage and volume consumed and this was converted into standard drinks (i.e., 12 oz of beer, 4 oz of wine, or 1.5 oz of spirits).

GWAS: Pre-genotyping QC of IASPSAD cases

DNA from all probands and affected siblings from the IASPSAD was quantified using both spectrophotometry and Quant-iT Picogreen (Life Technologies, Carlsbad, CA) technology. 200ng of each DNA sample was run on a 2% agarose gel to assess
degradation. Finally, all samples were sex-typed with sex chromosome-specific quantitative PCR (qPCR) using the Quantifiler Duo DNA Quantification Kit (Life Technologies, Carlsbad, CA). Only high molecular weight samples (bands > 800bp) with DNA concentrations ≥ 50ng/ul and with genotypic sex matching phenotypic sex were sent for genotyping.

**Genotyping, QC and Calling Algorithms**

815 samples from the IASPSAD, including eleven sets of parents to assess Mendelian errors, were sent to Vanderbilt University for genotyping on Affymetrix v6.0 arrays (Santa Clara, CA). 2000 Biobank controls from the Irish Blood Transfusion Service were also genotyped on Affymetrix v6.0 arrays at either the Affymetrix or Broad Institute core facilities. Arrays were called using the Affymetrix Birdseed v2 algorithm. Thirty-five samples were removed post-genotyping for CQC values <0.4, Nsp/Sty CQC values < 0 or discrepant genotypic sex calls. Arrays were then re-called using BEAGLECALL software (Browning, Yu 2009) in an effort to minimize site differences in genotyping. BEAGLECALL is an iterative process which starts with the fluorescent intensity calls and then incorporates haplotypic information from the sample with increasingly stringent calling filters. Subsequently, samples and SNPs underwent a thorough QC process. Samples were removed if the phenotypic and genotypic sex were discordant, they had excessive Mendelian errors, they were ancestry, heterozygosity or mean pi-hat outliers, their call rates were ≤ 98%, or if there was a DNA mix-up. SNPs were removed if HWE p-values < 1x10^{-6}, the minor allele frequency (MAF) < 1%, there were Mendelian errors, they were duplicates, their call rate < 98% or
if there was differential missingness between cases and controls. After QC, 676,736 SNPs were available for 710 probands and affected siblings and 1755 controls.

**Imputation**

Genotypes were imputed using IMPUTE2 (Howie, Donnelly et al. 2009) with the March 2012 1000 Genomes reference haplotype panel (1000 Genomes Project Consortium 2010). Probabilities were converted to dosages (based on number of minor alleles) using MACH2 (Scott, Mohlke et al. 2007). Post-imputation filtering removed SNPs with MAF < 1% and information content ≤ 3%, leaving approximately 8.2 million SNPs.

**Statistical Analyses**

Phenotype distribution was visualized using PASW 18 (SPSS Inc). Association analyses for MAX24 were performed within AD cases using ProbABEL (Aulchenko, Struchalin et al. 2010) in R 2.14.1 (R Core Team 2012). ProbABEL is unique in its ability to deal with related individuals, imputed genotypes and quantitative traits. To account for relatedness in the IASPSAD sample, ProbABEL calculates a kinship matrix. It then uses a two-step mixed model to perform linear regression. Sex and age at interview were included as covariates and residuals were used in the regression analysis. P-values were calculated using the Wald test. Based on the estimated 1.6 million independent tests in the 1000 Genomes dataset, genome-wide significance was set at 3.06x10^-8 (Li, Yeung et al. 2012). The false discovery rate (FDR) was calculated using the QVALUE package (Storey, Tibshirani 2003) in R (R Core Team 2012). Gene-
based analyses were performed using the Versatile Gene-Based test for Genome-wide Association Studies (VEGAS) program. The program assigns SNPs to genes, accounts for LD through a series of Monte Carlo simulations, and calculates an empirical gene-based test statistic for each gene (Liu, McRae et al. 2010). The authors suggest a Bonferroni-corrected significance threshold of $p<2.83 \times 10^{-6}$ (17,787 genes tested).

**Results**

Figure 5.1 shows the distribution of MAX24 values in the IASPSAD GWAS sample. MAX24 values range from 9 to 171 with a mean of 39.33 drinks in a twenty-four hour period. Though there is a wide distribution, overall the values are quite high, representative of the severely affected nature of the ascertained AD cases in the IASPSAD sample.

Figures 5.2 and 5.3 show the QQ plot and Manhattan plot, respectively, for MAX24 association analyses. There were twelve genome-wide significant SNPs, shown in Table 5.1. The top SNP, rs151338448 ($p=9.52 \times 10^{-12}$), is intronic to CCZ1B on chromosome 7. Rs117223671, rs188264442 and rs75214636 are intronic to SNTG1, PIK3R3 and ACOXL, respectively. Eight SNPs were intergenic and all clustered within approximately 100kb of SNTG1. Gene-based analysis yielded one gene, CREBBP on chromosome 16, that passed multiple testing correction (Table 5.2). Four other genes had FDR q-values less than 0.11 and all are located in a cluster on chromosome 1.
Figure 5.1. Distribution of MAX24 values in the IASPSAD (n=709).
Figure 5.2. **QQ plot for MAX24.** The red line represents distribution under the null hypothesis of no association. $\lambda = 0.999$.

Figure 5.3. **Manhattan plot for MAX24.** The red line represents genome-wide significance at $p \leq 3.06 \times 10^{-8}$.
Table 5.1. SNPs associated with MAX24 surviving multiple testing correction.

<table>
<thead>
<tr>
<th>Chr.</th>
<th>SNP</th>
<th>Position (bp)</th>
<th>MAF</th>
<th>Gene</th>
<th>p-Value</th>
<th>q-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>rs151338448</td>
<td>6862684</td>
<td>0.013</td>
<td>CCZ1B</td>
<td>9.52E-12</td>
<td>7.77E-05</td>
</tr>
<tr>
<td>8</td>
<td>rs191199258</td>
<td>50686000</td>
<td>0.015</td>
<td></td>
<td>6.29E-09</td>
<td>0.0110</td>
</tr>
<tr>
<td>8</td>
<td>rs117958407</td>
<td>50766599</td>
<td>0.016</td>
<td></td>
<td>6.44E-09</td>
<td>0.0110</td>
</tr>
<tr>
<td>8</td>
<td>rs74805105</td>
<td>50655986</td>
<td>0.016</td>
<td></td>
<td>6.62E-09</td>
<td>0.0110</td>
</tr>
<tr>
<td>8</td>
<td>rs76062094</td>
<td>50727424</td>
<td>0.015</td>
<td></td>
<td>6.81E-09</td>
<td>0.0110</td>
</tr>
<tr>
<td>8</td>
<td>rs117223671</td>
<td>50844885</td>
<td>0.014</td>
<td>SNTG1</td>
<td>1.28E-08</td>
<td>0.0167</td>
</tr>
<tr>
<td>1</td>
<td>rs188264442</td>
<td>46561551</td>
<td>0.010</td>
<td>PIK3R3</td>
<td>1.43E-08</td>
<td>0.0167</td>
</tr>
<tr>
<td>8</td>
<td>rs76936715</td>
<td>50806542</td>
<td>0.015</td>
<td></td>
<td>1.91E-08</td>
<td>0.0175</td>
</tr>
<tr>
<td>8</td>
<td>rs117015298</td>
<td>50806924</td>
<td>0.015</td>
<td></td>
<td>1.93E-08</td>
<td>0.0175</td>
</tr>
<tr>
<td>8</td>
<td>rs117455144</td>
<td>50780280</td>
<td>0.015</td>
<td></td>
<td>2.39E-08</td>
<td>0.0195</td>
</tr>
<tr>
<td>2</td>
<td>rs75214636</td>
<td>111602132</td>
<td>0.019</td>
<td>ACOXL</td>
<td>2.64E-08</td>
<td>0.0196</td>
</tr>
<tr>
<td>15</td>
<td>rs79326577</td>
<td>53190651</td>
<td>0.027</td>
<td></td>
<td>2.95E-08</td>
<td>0.0201</td>
</tr>
</tbody>
</table>

Table 5.2. Results for MAX24 gene-based association.

<table>
<thead>
<tr>
<th>Chr.</th>
<th>Gene</th>
<th>Start</th>
<th>Stop</th>
<th>No.SNPs</th>
<th>p-Value</th>
<th>q-Value</th>
<th>TopSNP</th>
<th>SNP p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>CREBBP</td>
<td>3715056</td>
<td>3870122</td>
<td>124</td>
<td>&lt;10^-6</td>
<td>0</td>
<td>rs129966</td>
<td>0.00038</td>
</tr>
<tr>
<td>1</td>
<td>MYCBP</td>
<td>39101222</td>
<td>39111637</td>
<td>104</td>
<td>2.10E-05</td>
<td>0.1055</td>
<td>rs4246511</td>
<td>1.64E-05</td>
</tr>
<tr>
<td>1</td>
<td>RRAGC</td>
<td>39077605</td>
<td>39097927</td>
<td>98</td>
<td>2.40E-05</td>
<td>0.1055</td>
<td>rs9439079</td>
<td>3.26E-05</td>
</tr>
<tr>
<td>1</td>
<td>RHBDL2</td>
<td>39124065</td>
<td>39180043</td>
<td>124</td>
<td>3.60E-05</td>
<td>0.1055</td>
<td>rs4246511</td>
<td>1.64E-05</td>
</tr>
<tr>
<td>1</td>
<td>GJA9</td>
<td>39112325</td>
<td>39119876</td>
<td>104</td>
<td>3.80E-05</td>
<td>0.1055</td>
<td>rs4246511</td>
<td>1.64E-05</td>
</tr>
</tbody>
</table>
Discussion

We performed a GWAS of MAX24 in our IASPSAD sample. There was a wide distribution of MAX24 values in the sample, but most were quite high with a mean of 39 and a maximum of 171 drinks in a twenty-four period. These values are comparable to the distribution seen in the COGA sample. Saccone et al reported an average of 29.5 drinks in a twenty-four hour period with a maximum value of 336 (Saccone, Kwon et al. 2000). In a non-ascertained sample, COGEND, the average maximum number of drinks in a twenty-four hour period was 16 overall, but 23 in males (Sherva, Rice et al. 2009).

The top SNP from the association analysis, rs151338448 (p=9.52x10^-12), is intronic to CCZ1B. CCZ1B, CCZ1 vacuolar protein trafficking and biogenesis associated homolog B (S. cerevisiae), and CCZ1, CCZ1 vacuolar protein trafficking and biogenesis associated homolog (S. cerevisiae), encode identical proteins, are located 873kb apart on chromosome 7, and are uncharacterized in humans but highly conserved in animals and plants. The yeast homolog of CCZ1 is integral to multiple vacuole processes (Wang, Stromhaug et al. 2002). In the mouse and Caenorhabditis elegans, the homolog for CCZ1 has been implicated in the processing and digestion of apoptotic cells (Kinchen, Ravichandran 2010, Nieto, Almendinger et al. 2010). There are no known reports linking CCZ1B or CCZ1 to alcohol consumption or AD.

Three other genes were implicated in the MAX24 SNP association analysis. SNTG1, γ1-syntrophin gene, belongs to the syntrophin family and the encoded protein is expressed only in neuronal cells (Piluso, Mirabella et al. 2000, Bashiardes, Veile et al. 2004). Disruption of SNTG1 has been documented in both scoliosis (Bashiardes, Veile
et al. 2004) and lung cancer in non-smokers (Job, Bernheim et al. 2010), but there are no references regarding this locus and MAX24 or AD. Although representing an indirect link, SNTG1 binds the neurotrophic protein ENO2, localizes it to the plasma membrane, and is crucial for its functioning (Hafner, Obermajer et al. 2010). ENO2 protein expression is decreased in heroin addicts compared to controls (Liao, Cheng et al. 2011). Similar pathways can be perturbed in addicts across drugs of abuse and perhaps this alternation is present in AD individuals as well.

**PIK3R3**, phosphoinositide-3-kinase regulatory subunit 3 (gamma), is in the lipid kinase family. These kinases have a plethora of functions within the cell (Koutros, Schumacher et al. 2010). Multiple sources report overexpression of PIK3R3 in gastric (Zhou, Chen et al. 2012), brain (Soroceanu, Kharbanda et al. 2007) and ovarian (Zhang, Huang et al. 2007) cancers and **PIK3R3** is associated with plaque formation in the brains of Alzheimer’s patients (Liang, Dunckley et al. 2008). **PIK3R3** is necessary for **IGF2** function (Soroceanu, Kharbanda et al. 2007). Igf2 protein expression is decreased in mouse embryos following alcohol exposure (Downing, Johnson et al. 2011) and IGF2 expression is statistically different in human children with prenatal alcohol exposure compared to unexposed children (Aros, Mills et al. 2011). Finally, the rat ortholog of **PIK3R3** was reported to have lower expression in the nucleus accumbens of alcohol-preferring rats self-administering ethanol compared to the same strain of rat self-administering saccharine (Rodd, Kimpel et al. 2008, Guo, Webb et al. 2009). This suggests that **PIK3R3** may be an ethanol responsive gene.

The final gene implicated in the MAX24 SNP association analysis is **ACOXL**. **ACOXL**, acyl-CoA oxidase-like, is uncharacterized in humans and has no known or
postulated role in maximum volume of alcohol consumed or AD. SNP variation in the gene is associated with risk of chronic lymphocytic leukemia in English (Di Bernardo, Crowther-Swanepoel et al. 2008), Chinese (Lan, Au et al. 2010), Spanish and Swedish (Crowther-Swanepoel, Mansouri et al. 2010) patients.

In addition to the genome-wide significant SNP results, gene-based analysis implicated five other genes in MAX24 in our sample (Table 2). CREBBP, CREB binding protein, on chromosome 16 was significantly associated after multiple testing correction. CREBPP binds CREB (cAMP response element binding protein) and facilitates its activation of cyclic adenosine monophosphate (cAMP) -responsive genes (Barnby, Abbott et al. 2005). Phosphorylation of CREBBP is integral in the well-documented ethanol-induced increase in cAMP signaling (Constantinescu, Wu et al. 2004) and has been implicated in a chromatin remodeling model of alcoholism (Pandey, Ugale et al. 2008). A human association study in heroin and alcohol addicts from India suggested that a SNP within CREBBP may be a risk factor for addiction (Kumar, Deb et al. 2011). In relation to alcohol consumption, Crebbp was identified in a quantitative trait loci (QTL) study of high and low alcohol-drinking rats (Bice, Liang et al. 2010).

Four other genes had FDR q-values less than 0.11 in the gene-based association analysis of MAX24: MYCBP, RRAGC, RHBDL2 and GJA9. RRAGC, Ras-related GTP binding C, has previous implications with alcoholic myopathy caused by heavy drinking. Briefly, ethanol increases RRAGC association with a key protein synthesis regulator, mechanistic target of rapamycin (MTOR). In so doing, ethanol causes a change in MTOR and decreases overall protein synthesis (Hong-Brown, Brown et al. 2012). RHBDL2, rhomboid, veinlet-like 2 (Drosophila), was reported to be
differentially expressed in the frontal cortex of alcoholics compared to non-alcoholics (Guo, Webb et al. 2009, Liu, Lewohl et al. 2006). The remaining genes have no known relationship with any alcohol-related phenotypes.

A GWAS of MAX24 yielded multiple genome-wide significant SNPs. Four SNPs were located in known loci. CCZIB, SNTG1, PIK3R3 and ACOXL represent new candidate loci for maximum alcohol consumed in twenty-four hours and possibly AD. Gene-based association analyses identified a loci, CREBBP, associated with MAX24 that has a rich history in the addiction literature. Although no SNPs within CREBBP reached genome-wide significance in the SNP association analysis, the aggregate association of multiple variants within the gene resulted in a significant gene-based test statistic. This suggests the power and utility of gene-based analyses. The IASPSAD is a severely affected, ascertained sample and therefore these results need to be explored in a population sample to determine applicability. The twelve genome-wide significant SNPs from the GWAS had low MAF, ranging from 1-3%. These results may need to be interpreted with caution. The IASPSAD sample size limits power to detect these smaller effect sizes and it is possible that a small number of genotypes could be artificially inflating the association results. However, the number of SNPs clustering on chromosome 8 in the genome-wide significant results argues against that, at least in one case. Replication in other human datasets would strengthen the association results and molecular characterization is necessary to untangle the role of each associated loci in the MAX24 phenotype in our sample.
Future Directions

These results represent the initial GWAS analysis of the MAX24 phenotype in the IASPSAD sample. There are multiple additional studies to be undertaken. As noted earlier, values for the MAX24 phenotype are quite high and may represent an over-inflation in reported values. One method to account for this inflation is transformation of the data, prior to association analyses, as suggested in Saccone et al (Saccone, Kwon et al. 2000). A thorough analysis of SNPs in linkage disequilibrium (LD) with the top SNPs is integral to analysis of the data. Top SNPs in LD with other associated SNPs are more indicative of a true association. Genes associated with MAX24 in our sample will be followed-up for association with AD as well. As mentioned previously, replication will be attempted in the COGA sample to strengthen the association results. Candidate genes from previous studies of MAX24 and the linkage regions implicated in the IASPSAD sample will be examined in the data, possibly by weighting these results and re-analyzing the data. For significantly associated loci, messenger RNA (mRNA) differences will be assessed in postmortem brain samples from chronic alcoholics and controls. Also, Association LIst Go AnnoTatOR (ALIGATOR) will be run in the dataset to identify pathways and Gene Ontology groups that are overrepresented (Holmans, Green et al. 2009).

Finally, a major strength of physiological phenotypes in human research is their easier application to model organisms. The results presented here-in have not yet been explored in-depth in the model organism literature. It may also be worthwhile to examine genes of interest in model organisms represented within the VCU Alcohol Research Center.
Examination of the Shared Variation Between Alcohol Dependence, Major Depression and Neuroticism

Introduction

Previous chapters have examined genetic risk factors for alcohol dependence (AD) and multiple alcohol-related phenotypes (ARPs) including alcohol craving and initial sensitivity. Examining these physiological traits which correlate with risk is integral to understanding the phenotypic complexity of AD. However, it is also prudent to consider the high levels of comorbidity seen with AD. A wide range of common psychiatric and substance use disorders are seen in AD individuals. In concert with ARPs, it is only by understanding the relationship between AD and comorbid phenotypes, and the genetic architecture contributing to each, that researchers can begin to untangle the genetics of alcohol use disorders. This chapter presents ongoing work in examining the relationship between AD and comorbid internalizing disorders.

Comorbidity is the norm rather than the exception in common psychiatric (e.g. major depression) and substance use disorders (e.g. alcohol dependence) (Kessler, Nelson et al. 1996, Regier, Farmer et al. 1990). Internalizing disorders, characterized by negative mood states, include major depression, dysthymia, generalized anxiety disorder and phobias. Externalizing disorders, characterized by acting out and behavioral disinhibition, include substance use disorders, adult antisocial personality disorder and conduct disorder. AD correlates most strongly with the externalizing disorders, but is significantly correlated with internalizing disorders as well (Kessler,
Nelson et al. 1996, Kessler, Crum et al. 1997, Kessler, Chiu et al. 2005, Regier, Farmer et al. 1990, Dawson, Goldstein et al. 2010). Furthermore, a recent examination in the National Epidemiological Survey on Alcohol and Related Conditions (NESARC) showed that one-third of alcohol-related problems and outcomes were explained by the shared variation seen amongst internalizing disorders (Kushner, Wall et al. 2012). Given the high levels of comorbidity seen in these disorders, it is not unreasonable to hypothesize they share genetic liability. A seminal paper by Kendler et al examined externalizing and internalizing disorders in a Virginia twin sample. Within their data, one common genetic factor loaded onto the externalizing disorders, including AD, and, though the architecture was more complex, one common genetic factor loaded onto the internalizing disorders (Kendler, Heath et al. 1992). Although AD was related most strongly to the externalizing factor, there was a significant, though smaller, loading of the common internalizing genetic factor onto AD. A final piece of evidence supporting the relationship between AD and both externalizing and internalizing disorders is found in the phenotypic differences among alcoholics.

Numerous studies have examined typologies of AD individuals. Perhaps the most well-known classification system is that of Cloninger who identified two types of alcoholics: Type I alcoholics had a milder form of alcoholism, a later onset of the disease and were more sensitive to the environment whereas Type II alcoholics had a more severe form of alcoholism, were typically male, and were more likely to have had criminal problems (Cloninger, Bohman et al. 1981, Saunders 1982). Using Cloninger’s typology, later studies reported that Type II alcoholics were also more likely to exhibit externalizing behaviors like substance abuse and antisocial personality disorder.
(Hallman, von Knorring et al. 1996, von Knorring, Bohman et al. 1985). Other well-known classification systems have found similar phenotypic stratifications. Zucker’s model divided alcoholics into four groups, two of which were antisocial alcoholism and negative affect alcoholism (Zucker 1986). Windle et al also defined four types of alcoholism: mild, polydrug, negative affect and antisocial (Windle, Scheidt 2004). Both Lesch and Del Boca’s typologies included subgroups with depression and anxiety and others with behavioral disturbances and antisocial personality. Though there have been numerous classifications of alcoholics, and much debate as to the best system, a general theme is that there exist at least two types of alcoholics. One group, which could be called the “internalizers”, drink to alleviate negative mood states whereas another group, the “externalizers”, drink as part of their overall behavioral disinhibition. Although the literature suggests that AD is more strongly correlated with externalizing disorders, due to the high levels of comorbidity seen between AD and internalizing disorders, the known differences in types of alcoholics and the documented shared genetic factors, it is of interest to examine genetic risk loci that are associated with risk for this broader phenotypic liability.

This chapter focuses on understanding the shared variation between AD, major depression (MD) and neuroticism. AD has been discussed extensively in the Introduction. MD is defined by the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) as a depressed mood or loss of interest for at least two weeks and at least four of the following nine symptoms during that same period: depressed mood, anhedonia, weight changes, sleep changes, motor disturbances, fatigue, feelings of worthlessness or guilt, difficulty concentrating and
recurrent thoughts of death and suicide (American Psychiatric Association 2000). Neuroticism is a personality trait characterized by emotional distress. Feelings of nervousness, guilt and depression are felt by persons with high levels of neuroticism as are emotional instability and low self-esteem (Wray, Middeldorp et al. 2008).

The comorbidity seen between AD and MD is well established. Petrakis et al reported individuals with AD in the Epidemiologic Catchment Area (ECA) study and National Comorbidity Survey (NCS) were 3.9 times more likely in the past year to have had MD in comparison to those without AD (Petrakis, Gonzalez et al. 2002, Regier, Farmer et al. 1990, Kessler, Nelson et al. 1996). Similarly, the National Longitudinal Alcohol Epidemiological Survey (NLAES) found that individuals with AD were 4.2 times more likely in the previous year to have reported MD (Hasin, Grant 2002). In the NESARC, individuals with AD were 3.7 times more likely to have MD in the previous year (Grant, Dawson et al. 2004). Although there have been numerous studies looking for genetic risk factors of AD and MD independently, fewer studies have looked at the comorbid phenotype. Nurnberger et al found linkage to chromosome 1 for individuals with either AD or MD, suggesting that there were genetic risk factors in this region that could affect risk for both phenotypes (Nurnberger, Foroud et al. 2002). Candidate gene studies have found association between CHRM2 (Wang, Hinrichs et al. 2004) and SLC6A4 (Marques, Hutz et al. 2006) in both AD and MD. Interestingly, a study in a Han Chinese sample found that the DRD2 gene was associated with subjects endorsing AD, MD and anxiety but not in subjects with only AD, MD or anxiety (Huang, Lin et al. 2004). Kertes et al examined an array of candidate genes in comorbid individuals and found evidence of association with CRHBP, OPRM1 and GABRB1 (Kertes, Kalsi et al.
The only genome-wide association study (GWAS) of AD and MD was recently published by Edwards et al. No single nucleotide polymorphisms (SNPs) were genome-wide significant, but pathway analyses revealed enrichment of glutamatergic function genes (Edwards, Aliev et al. 2012).

Neuroticism is strongly correlated with the internalizing disorder spectrum (Bienvenu, Brown et al. 2001) and high levels of neuroticism have been reported to predict higher numbers of internalizing symptoms (Andrews, Stewart et al. 1990). A Virginia twin study by Hettema et al found that a common genetic factor for the internalizing disorders also significantly loaded onto neuroticism (Hettema, Neale et al. 2006). The literature also reports that neuroticism is correlated with future MD episodes (Hirschfeld, Klerman et al. 1989, Boyce, Parker et al. 1991) and AD. An English twin study found higher neuroticism scores in drinking versus non-drinking twins (Mullan, Gurling et al. 1986) and a separate group found high levels of neuroticism were correlated with an increased risk for AD (Loukas, Krull et al. 2000). Numerous studies have used linkage, candidate gene association studies and GWAS to investigate genetic risk factors for AD and neuroticism separately, but at least two studies have assessed genetic risk factors for both phenotypes within the same sample. Rs7590720, in the PECR gene, is associated with both AD and neuroticism in a German sample (Grabe, Mahler et al. 2011) and substance dependent African American females showed association between ADH1A and neuroticism (Zuo, Gelernter et al. 2010).

Within this chapter, preliminary results from the IASPSAD and COGA samples are presented which examine the relationship between AD, MD and neuroticism. Factor
analysis is used to identify shared variation and preliminary invariance modeling is explored to compare latent mean differences between groups.

Methods

Samples: Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD)

A detailed description of the sample is provided elsewhere (Prescott, Sullivan et al. 2006). Briefly, probands were ascertained in clinical treatment facilities and hospitals in the Republic of Ireland and Northern Ireland between 1998 and 2002. Probands were interviewed using an adapted version of the Semi-Structured Assessment of the Genetics of Alcoholism (SSAGA) and diagnosed with AD using DSM-IV criteria. The sample is ethnically homogenous; a stipulation for inclusion in the study required each proband have all 4 grandparents born in Ireland, Northern Ireland, Scotland, Wales or England. From the original linkage sample, 710 probands and affected siblings are used in this study.

Samples: Collaborative Study on the Genetics of Alcoholism (COGA)

Probands for the COGA sample, details of which are described elsewhere (Dick, Plunkett et al. 2006), were recruited throughout the United States in 6 geographic locations from alcohol treatment programs. Large families were preferentially ascertained with a proband required to have 2 or more family members in the COGA sampling area and preferably sibships greater than 3. All individuals were given the SSAGA interview and AD diagnosed using DSM-IV criteria. From the larger COGA
sample, 847 genetically independent AD cases and 552 controls of European ancestry were used.

Phenotypes

AD and MD were measured identically in both samples, using the SSAGA interview. Symptom counts for each were used for factor analysis. Two different, but related, measures of neuroticism were used in the IASPSAD and COGA samples. The IASPSAD sample used the Eysenck Personality Questionnaire-Revised-Shortened (EPQ-RS) (Eysenck, Eysenck et al. 1985). The EPQ-RS is a yes/no self-report form with twelve questions regarding neuroticism. The COGA sample used the NEO Five-Factor Inventory (NEO-FFI) (Costa Jr., McCrae 1992). The NEO-FFI has twelve items regarding neuroticism, each using a five point Likert scale. In each sample, sum scores for neuroticism were used for factor analysis. N=706 AD cases, 698 individuals with MD symptoms and 677 people with neuroticism scores in the IASPSAD sample. N= 844 AD cases, 844 individuals with MD symptoms and 567 people with neuroticism scores in the COGA sample.

Factor analysis

All analyses described within this section were done separately, but identically, in each sample unless otherwise noted. All three variables (AD symptom count, MD symptom count and neuroticism sum scores) were non-normal and therefore ordinalized in PASW18 (SPSS Inc.) prior to factor analysis. Factor analysis was performed in Mplus (Muthén, Muthén 1998-2011) using the weighted least squares means and
variances adjusted (WLSMV) estimator. The WLSMV is a two stage process and is well suited for non-normal and ordinal data (Muthén, du Toit et al. accepted for publication). It operates by first estimating a pairwise variable by variable polychoric matrix from which it creates an unobserved latent variable, for each observed variable, that is assumed to be normally distributed. From these additional latent response variables, the WLSMV fits the model. Males and females were run together and sample identification (IASPSAD or COGA) was used as a grouping variable. As the IASPSAD sample included siblings, the cluster option was used. In the COGA sample, each sample was given its own family number.

*Invariance modeling factor analysis*

To calibrate factor scores between groups, invariance modeling was performed in Mplus in addition to all previously discussed coding options. In the baseline, saturated model, all metrics were unrestrained. In the full invariance model, factor loadings and thresholds were held invariant between the IASPSAD and COGA samples. In the partial invariance models, one variable’s factor loadings and thresholds were allowed to be variant between samples. Using the DIFFTEST metric in Mplus, a test statistic was created for each nested model, comparing fit to the baseline model. Model fit was also assessed using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error Approximation (RMSEA). CFI and TLI values between 0.90 and 0.95 are considered acceptable while values between 0.95 and 1 are considered good. RMSEA values ≤0.05 are considered good and ≤0.10 acceptable (Levine, Petrides et al. 2005, Hu, Bentler 1995)
Statistical Analyses

Phenotype distribution visualization and correlations were performed using PASW 18 (SPSS Inc).

Results

In the IASPSAD, the average AD symptom count was 6.43 and the average MD symptom count 5.88. Neuroticism sum scores ranged from 0 to 12 with a mean of 8.11.

In the COGA sample, 5.62 was the average AD symptom count. The average MD symptom count was 4.97. Neuroticism sum scores ranged from 26 to 74 with a mean of 58. All phenotypes were significantly correlated with each other (Table 6.1). All variable distributions showed a non-normal distribution (Figure 6.1). To combat this non-normality, values were ordinalized: AD into three groups, MD into four and neuroticism into five.

Table 6.1. Correlations between phenotypes in the IASPSAD and COGA.

<table>
<thead>
<tr>
<th>IASPSAD</th>
<th>ADsx</th>
<th>MDsx</th>
<th>Nss</th>
<th>COGA</th>
<th>ADsx</th>
<th>MDsx</th>
<th>Nss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADsx</td>
<td>1</td>
<td></td>
<td></td>
<td>ADsx</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDsx</td>
<td>0.252</td>
<td>1</td>
<td></td>
<td>MDsx</td>
<td>0.204</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Nss</td>
<td>0.304</td>
<td>0.299</td>
<td>1</td>
<td>Nss</td>
<td>0.184</td>
<td>0.339</td>
<td>1</td>
</tr>
</tbody>
</table>

All correlations were significant at the 0.01 level (2-tailed). ADsx = AD symptom count, MDsx = MD symptom count and Nss = neuroticism sum scores.
Factor analysis was performed in each sample using the ordinalized AD, MD and neuroticism phenotypes. The factor loadings were similar across samples (Table 6.2). MD symptom count loaded very strongly onto the factor (0.7 and 0.717 in the IASPSAD and COGA samples, respectively). AD symptom count loaded moderately well onto the factor (0.544 and 0.420 in the IASPSAD and COGA samples, respectively) and neuroticism loaded fairly well onto the factor (0.554 and 0.495 in the IASPSAD and COGA samples, respectively).

Table 6.2. Factor loadings on the common factor.

<table>
<thead>
<tr>
<th></th>
<th>IASPSAD</th>
<th>COGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADsx</td>
<td>0.544</td>
<td>0.420</td>
</tr>
<tr>
<td>MDsx</td>
<td>0.702</td>
<td>0.717</td>
</tr>
<tr>
<td>Nss</td>
<td>0.554</td>
<td>0.495</td>
</tr>
</tbody>
</table>

ADsx= AD symptom counts, MDsx = MD symptom counts, and Nss = neuroticism sum scores.
Alcohol Dependence symptoms (3-7)

Major Depression symptoms (0-9)

Neuroticism sum scores (varies)

Figure 6.1. Phenotype distributions in the IASPSAD and COGA.
Discussion

This chapter presents preliminary data from a factor analysis of AD, MD and neuroticism in the IASPSAD and COGA samples. Neuroticism was measured with the EPQ-RS in the IASPASD sample and the NEO-FFI in the COGA sample. Both forms probe feelings of anxiety and depression. However, the NEO-FFI also asks about angry hostility, self-consciousness, impulsiveness and vulnerability. A review of the literature indicates that multiple measures of neuroticism can be appropriately combined within a study. A general population cohort from Scotland (n=207) was administered both interviews. The phenotypic correlation was 0.89 (Gow, 2005). In an Australian twin study, individuals were given the EPQ-RS in 1989 and NEO-FFI in 2002. The genetic correlation was 0.88 (Wray, 2008). All data in the literature suggests that there is a positive phenotypic and genetic correlation between the two indices of neuroticism used in the study.

The phenotypic distributions of AD, MD and neuroticism were all non-normal. AD and neuroticism were skewed to the right as would be expected in a sample ascertained for AD. The distribution of MD symptoms, in each sample, showed a large number of individuals with no symptoms. Among the rest of the subjects, the distribution of symptoms was skewed to the right. This is likely an artifact of the interview process: if an individual failed to meet criteria during two screening questions, no further symptoms of MD were queried. This resulted in a large number of individuals with no MD symptoms. The remainder of the MD distribution is due to the severely affected, ascertained nature of the samples used in this study.
Factor analysis was used to examine the shared variation between AD, MD and neuroticism. Simply put, factor analysis is a data reduction technique that attempts to find patterns in the data; groups of data that function like each other (Fabrigar, Wegener et al. Psychological Methods). A common factor was derived, explaining a large amount of the variance for MD and a smaller amount of the variance for AD and neuroticism. The nature of the factor loadings make it unclear if the common factor is adequately measuring the shared variation between the disorders. Especially in the COGA sample, the factor loadings for AD and neuroticism (0.420 and 0.495, respectively) are quite modest. It is hypothesized that the current analysis is not the most effective way to analyze the data. As will be discussed in the Future Directions section, an item level factor analysis will shed light on the underlying structure of the variables in each sample and inform which statistical method is best suited for further research.

There are several other aspects of this study that warrant further discussion. First, both samples contain comorbid cases whose MD and AD did not occur independently of each other. Therefore these people do not meet the classical criteria for MD. Edwards et al, defines these individuals as having a depressive syndrome rather than MD (Edwards, Aliev et al. 2012). Secondly, the temporal relationship between AD and MD in each sample is not addressed. In the literature, the nature of the genetic liability for internalizing disorders and AD is still unclear. It remains to be determined if certain loci predispose individuals to both AD and MD or rather if certain genes predispose individuals to one which then leads to the other. It is also possible that both hypotheses are correct and some genetic risk factors do predispose persons
to a broader comorbidity and other risk factors are disorder specific. The literature contains studies that have found MD indicative of future AD (Kuo, Gardner et al. 2006, Sihvola, Rose et al. 2008) and others that suggest AD leads to MD (Fergusson, Boden et al. 2011, Schuckit, Smith et al. 2007). In the IASPSAD, Kertes et al reported that the relationship was murky (Kertes, Kalsi et al. 2011) and in the COGA sample, Edwards et al reported that in a majority of the females, MD preceded AD (Edwards, Aliev et al. 2012). It will be of use to examine this in the particular sample subsets used to determine the nature of the relationships. It is unlikely we will have power to split up the sample into different groups, but it will be useful to inform our analyses.

FUTURE DIRECTIONS

In order to inform comparisons between the IASPSAD and COGA, and to calibrate factor scores between the samples (Milfont, Fischer 2010), preliminary invariance modeling was performed. In the baseline model, all metrics were unrestrained. (This is the model presented in the chapter results.) In the full invariance model, factor loadings and thresholds were held invariant between the IASPSAD and COGA samples. In the partial invariance models, one variable’s factor loadings and thresholds were allowed to be variant (“free”) between samples while the rest remained invariant. The model fitting results are reported in Table 6.3 where a significant p-value corresponds to a model fitting significantly worse than the baseline model. The full invariance model fit was poor. A partial invariance model with AD symptom count unrestrained, but MD and neuroticism invariant, was not significantly different than the baseline model. Therefore,
it was chosen as the best fitting model. The factor loadings under this model are presented in Table 6.4.

Table 6.3. Invariance modeling results.

<table>
<thead>
<tr>
<th></th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>Chi-square</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full invariance</td>
<td>0.607</td>
<td>0.633</td>
<td>0.156</td>
<td>139.347</td>
<td>7</td>
<td>0.0000</td>
</tr>
<tr>
<td>Partial_ADfree</td>
<td>0.984</td>
<td>0.984</td>
<td>0.035</td>
<td>11.605</td>
<td>6</td>
<td>0.0714</td>
</tr>
<tr>
<td>Partial_MDfree</td>
<td>0.672</td>
<td>0.606</td>
<td>0.169</td>
<td>115.968</td>
<td>5</td>
<td>0.0000</td>
</tr>
<tr>
<td>Partial_Nfree</td>
<td>0.732</td>
<td>0.597</td>
<td>0.171</td>
<td>94.490</td>
<td>4</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

N = neuroticism.

Table 6.4. Factor loadings using the partial invariance model with AD variant.

<table>
<thead>
<tr>
<th></th>
<th>IASPSAD</th>
<th>COGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADsx</td>
<td>0.556</td>
<td>0.386</td>
</tr>
<tr>
<td>MDsx</td>
<td>0.804</td>
<td>0.804</td>
</tr>
<tr>
<td>Nss</td>
<td>0.477</td>
<td>0.477</td>
</tr>
</tbody>
</table>

ADsx = AD symptom counts, MDsx = MD symptom counts, and Nss = neuroticism sum scores.
Several problems are evident. The best model was a partial invariance model with MD and neuroticism invariant and AD loadings and thresholds allowed to vary between samples. Traditionally, researchers have tested the full invariance model and if the fit was poor, no further tests were done. The idea of partial invariance modeling was developed by Muthen et al in 1981 (Muthen, Christoffersson 1981) and was subsequently shown to allow for comparison of variables between groups (Byrne, Shavelson et al. 1989). Partial invariance modeling has gained acceptance in the field, however the number of variable items allowed in the model is still unresolved (Millsap, Kwok 2004). Bryne et al reported that a majority of the items in the factor analysis must be invariant (Byrne, Shavelson et al. 1989) while Reise et al reported that one-half of the items should be invariant (Reise, Widaman et al. 1993). Recent work has suggested that model size (the number of items being tested) is also integral to the validity and effectiveness of partial invariance modeling (Donahue 2006). The partial invariance model presented in this chapter was comprised of only three items and so may be psychometrically invalid. Furthermore, even when using the partial invariance model, the factor loadings are again troublesome. In the COGA sample, AD symptom count loads very weakly (0.386) onto the common factor suggesting that factor is not capturing a substantial amount of variance in the trait.

Current research is ongoing to address these issues. Instead of using symptom counts and sum scores, an item level factor analysis will be done using invariance modeling and grouping by sex and sample (IASPSAD vs. COGA). The item level data for neuroticism is unavailable in the COGA sample, so the analysis will focus on AD and MD, only. Item level analysis will allow for examination of patterns in the data and a
more direct, concrete knowledge of the phenotype specifically being measured in the analysis. Once the best-fitting model is chosen, GWAS, meta-analysis and gene-based association analyses can be re-run in both samples. Importantly, pathway-based analysis will be performed to see if groups of genes associated with common functionality are replicated between samples.

In conclusion, this chapter presents preliminary data from work done in two ascertained samples: the IASPSAD and COGA. Factor analysis was used to examine shared variation between AD, MD and neuroticism. Work continues to improve the factor analysis model for internalizing disorders and AD. Previous work has been done in the IASPSAD looking at externalizing disorders and AD (Hack et al, unpublished), while previous chapters have examined biochemical risk factors for AD. When all analyses are complete, there will be a more global picture of risk pathways for AD in the IASPSAD sample.
CHAPTER SEVEN

Global Conclusions and Future Directions

This dissertation has presented multiple alternative strategies to identifying genetic risk factors for Alcohol Dependence (AD) and alcohol-related phenotypes. Chapter Two reports association between variants in GPSM1 and craving-related behaviors, suggesting that the model organism literature provides a targeted, improved method for selection of candidate genes in addiction research. Chapter Three details significant association between two loci on chromosome 7, KIAA1324L and GRM3, and AD in addition to significant gene expression differences in GRM3 between chronic alcoholics and controls. This region on chromosome 7 was identified in a pooled genome-wide association study (GWAS), illustrating the utility of GWAS studies to identify new risk loci. The remainder of the dissertation is focused on genome-wide analyses of physiological traits and comorbid phenotypes correlated with risk for AD within the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD). Chapter Four examines initial sensitivity, Chapter Five, maximum alcohol consumption in twenty-four hours and Chapter Six, comorbidity between AD, major depression and neuroticism.

Three overall themes emerge from this dissertation regarding phenotypic complexity and GWAS methodology. First, in the two GWAS performed in this dissertation, there are no overlapping single nucleotide polymorphisms (SNPs) or loci in the top results. Furthermore, there are no overlapping signals in the top results between any of these analyses and an unpublished GWAS of AD in the IASPSAD sample (Hack et al, unpublished results). This suggests, at least within the IASPSAD, that different
genetic risk loci predispose individuals to different alcohol-related phenotypes. Also, the
diagnostic phenotype of AD does not identify these risk factors. Therefore, it is
imperative that alcohol researchers adequately define their phenotypes and take into
account comorbidity within their samples. Proper measurement of phenotypes of
interest may necessitate advanced structural equation modeling, as discussed in
Chapter Six.

Secondly, results from the initial sensitivity (Chapter Four) GWAS suggest loci
affecting risk for this phenotype are involved in basic, ubiquitous cell signaling.
Genome-wide significant variants were located within genes whose functions included
degradation of proteins, transport of molecules and cell-to-cell connections. If it is true,
these results suggest a very complex molecular biology of the phenotype involving
numerous loci. Many of the loci may have other loci in the genome with compensatory
roles. This complexity may help explain the missing heritability described in the
literature and discussed below.

Lastly, the two GWAS analyses presented in this dissertation were successful in
spite of a relatively small sample size (n=710 AD cases). Although further examination
of association signals in linkage disequilibrium (LD) with top SNPs and replication in
other datasets is still needed, the GWAS of initial sensitivity identified nineteen genome-
wide significant SNPs and maximum consumption, twelve genome-wide significant
SNPs. Before and after genotyping, IASPSAD samples were subjected to a thorough
QC process. Additionally, samples were called using both fluorescence intensity (the
common method) and LD information. The rigorous QC and calling process is
hypothesized to have reduced noise within the sample, through elimination of poorly performing samples and SNPs, and resulted in significant, high quality results.

GWAS have identified a number of potential genetic risk factors for AD and related phenotypes. However, there are two conspicuous limitations: a lack of replication and the issue of ‘missing heritability’: identified variants explaining a minimal amount of variance in the trait. An often used example is that of height. Unlike alcohol dependence, and many other complex disorders, height is an unambiguous phenotype that can be reliably and directly measured. Despite this, as of 2011 it was estimated that 180 loci identified in GWAS for height explained only 12% of the variance in the trait (Lander 2011). Currently, collaborations are being planned to perform GWAS on alcohol-related phenotypes in large samples of thousands of individuals (Agrawal, Freedman et al. 2012). This increase in power should enable better identification of the many genetic risk loci hypothesized to affect risk for these phenotypes (Spencer, Su et al. 2009). However, as this dissertation suggests, it will be imperative that researchers thoroughly QC their data beforehand and, if re-calling of genotypes is attempted, that site differences be identified and addressed.

Alternatively, the problem of missing heritability could be the result of a flaw in the accepted model of complex inheritance. As discussed in the Introduction, the GWAS era is based on the common disease - common variant (CDCV) hypothesis wherein a large number of loci of small effect size contribute to risk for AD. In contrast, the common disease – rare variant (CDRV) model hypothesizes that a few rare variants of large effect contribute to the risk for AD (Rowe, Tenesa 2012). It seems plausible, as hypothesized by some researchers (e.g. (Gibson 2012)) that both models operate
simultaneously. Rare variation has not been well studied in AD but the advancing technology and decreasing cost of exome sequencing should result in a wealth of large datasets in the next few years.

Finally, gene expression studies can be used to either identify differentially expressed loci in alcoholics compared to controls or to study a candidate loci’s involvement in a phenotype of interest. Human postmortem brain tissue banks and array technology have allowed for large-scale examinations of messenger RNA (mRNA) expression (Crabbe 2008). Brain regions studied for differences between alcoholics and controls include the frontal cortex (Liu, Lewohl et al. 2006, Lewohl, Wang et al. 2000), nucleus accumbens (Flatscher-Bader, van der Brug et al. 2005), hippocampus (Enoch, Zhou et al. 2012) and amygdala (Ponomarev, Wang et al. 2012). These studies have typically found large numbers of differentially expressed genes and implicated major signaling pathways like GABAergic and glutamatergic neurotransmission. By using convergent lines of evidence such as GWAS data, exome sequencing and gene expression, researchers will be able to comprehensively investigate genetic risk factors for AD and related phenotypes.


BACKSTROM, P. and HYYTIA, P., 2005. Suppression of alcohol self-administration and cue-induced reinstatement of alcohol seeking by the mGlu2/3 receptor agonist LY379268 and the mGlu8 receptor agonist (S)-3,4-DCPG. *European journal of pharmacology,* 528(1-3), pp. 110-118.


HIRSCHHORN, J.N., 2005. Genetic approaches to studying common diseases and complex traits. *Pediatric research, 57*(5 Pt 2), pp. 74R-77R.


KOUTROS, S., SCHUMACHER, F.R., HAYES, R.B., MA, J., HUANG, W.Y., ALBANES, D., CANZIAN, F.,
CHANOCK, S.J., CRAWFORD, E.D., DIVER, W.R., FEIGELSON, H.S., GIOVANUCCI, E., HAIMAN,
C.A., HENDERSON, B.E., HUNTER, D.J., KAAKS, R., KOLONEL, L.N., KRAFT, P., LE MARCHAND, L.,
RIBOLI, E., SIDIQI, A., STAMPFER, M.J., STRAM, D.O., THOMAS, G., TRAVIS, R.C., THUN, M.J.,
and risk of prostate cancer. Cancer research, 70(6), pp. 2389-2396.

KUMAR, D., DEB, I., CHAKRABORTY, J., MUKHOPADHYAY, S. and DAS, S., 2011. A polymorphism of
the CREB binding protein (CREBBP) gene is a risk factor for addiction. Brain research, 1406, pp. 59-64.

KUMAR, S., PORCU, P., WERNER, D.F., MATTHEWS, D.B., DIAZ-GRANADOS, J.L., HELFAND, R.S.
and MORROW, A.L., 2009. The role of GABA(A) receptors in the acute and chronic effects of ethanol: a

KUO, P.H., GARDNER, C.O., KENDLER, K.S. and PRESCOTT, C.A., 2006. The temporal relationship of
the onsets of alcohol dependence and major depression: using a genetically informative study design.
Psychological medicine, 36(8), pp. 1153-1162.

KUO, P.H., KALSI, G., PRESCOTT, C.A., HODGKINSON, C.A., GOLDMAN, D., ALEXANDER, J., VAN
DEN OORD, E.J., CHEN, X., SULLIVAN, P.F., PATTERTSON, D.G., WALSH, D., KENDLER, K.S. and
RILEY, B.P., 2009. Associations of glutamate decarboxylase genes with initial sensitivity and age-at-
onset of alcohol dependence in the Irish Affected Sib Pair Study of Alcohol Dependence. Drug and

KUO, P.H., NEALE, M.C., RILEY, B.P., WEBB, B.T., SULLIVAN, P.F., VITTUM, J., PATTERTSON, D.G.,
Identification of susceptibility loci for alcohol-related traits in the Irish Affected Sib Pair Study of Alcohol
Dependence. Alcoholism, Clinical and Experimental Research, 30(11), pp. 1807-1816.

KUSHNER, M.G., WALL, M.M., KRUEGER, R.F., SHER, K.J., MAURER, E., THURAS, P. and LEE, S.,
2012. Alcohol dependence is related to overall internalizing psychopathology load rather than to particular
internalizing disorders: evidence from a national sample. Alcoholism, Clinical and Experimental Research,
36(2), pp. 325-331.

LAN, Q., AU, W.Y., CHANOCK, S., TSE, J., WONG, K.F., SHEN, M., SIU, L.P., YUENGER, J., YEAGER,


LANDER, E. and KRUGLYAK, L., 1995. Genetic dissection of complex traits: guidelines for interpreting

LANDER, E.S., LINTON, L.M., BIRREN, B., NUSBAUM, C., ZODY, M.C., BALDWIN, J., DEVON, K.,
DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A.,
HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J.,
MESIROV, J.P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R.,
SHERIDAN, A., SOUZNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A.,
WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J.,
CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I.,
DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A.,
JONES, M., LLOYD, C., McMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J.C.,


Sidhpura, N., Weiss, F. and Martin-Fardon, R., 2010. Effects of the mGlu2/3 agonist LY379268 and the mGlu5 antagonist MTEP on ethanol seeking and reinforcement are differentially altered in rats with a history of ethanol dependence. Biological psychiatry, 67(9), pp. 804-811.


SLUTSKE, W.S., TRUE, W.R., SCHERRER, J.F., HEATH, A.C., BUCHOLZ, K.K., EISEN, S.A.,
GOLDBERG, J., LYONS, M.J. and TSUANG, M.T., 1999. The heritability of alcoholism symptoms:
"indicators of genetic and environmental influence in alcohol-dependent individuals" revisited. Alcoholism, 
Clinical and Experimental Research, 23(5), pp. 759-769.

mitochondrial, ubiquitin and signaling systems abnormalities in postmortem brains from subjects with a
history of alcohol abuse or dependence. Journal of neuroscience research, 72(6), pp. 756-767.

SOLOMON, R.L., 1980. The opponent-process theory of acquired motivation: the costs of pleasure and

SONG, J., KOLLER, D.L., FORoud, T., CARR, K., ZHAO, J., RICE, J., NURNBERGER, J.I., Jr,
of GABA(A) receptors and alcohol dependence and the effects of genetic imprinting. American journal
of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of

SOROCEANU, L., KHARBANDA, S., CHEN, R., SORIANO, R.H., ALDAPE, K., MISRA, A., ZHA, J.,
Identification of IGf2 signaling through phosphoinositide-3-kinase regulatory subunit 3 as a growth-
promoting axis in glioblastoma. Proceedings of the National Academy of Sciences of the United States of
America, 104(9), pp. 3466-3471.

SOYKA, M., PREUSS, U.W., HESSELBROCK, V., ZILL, P., KOLLER, G. and BONDY, B., 2008. GABA-
A2 receptor subunit gene (GABRA2) polymorphisms and risk for alcohol dependence. Journal of
psychiatric research, 42(3), pp. 184-191.

e1000477.

of the National Academy of Sciences of the United States of America, 100(16), pp. 9440-9445.

SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V.K., MUKHERJEE, S., EBERT, B.L., GILLETTE, M.A.,
enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.
Proceedings of the National Academy of Sciences of the United States of America, 102(43), pp. 15545-
15550.

TAKEUCHI, F., ISONO, M., NABIKA, T., KATSUYA, T., SUGIYAMA, T., YAMAGUCHI, S., KOBAYASHI,
locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese

TEAGUE, B., WATERMAN, M.S., GOLDSTEIN, S., POTAMOUSIS, K., ZHOU, S., RESLEWIC, S.,
SARKAR, D., VALOUEV, A., CHURAS, C., KIDD, J.M., KOHN, S., RUNNHEIM, R., LAMERS, C.,
FORREST, D., NEWTON, M.A., EICHLER, E.E., KENT-FIRST, M., SURLT, U., LIVNY, M. and
Proceedings of the National Academy of Sciences of the United States of America, 107(24), pp. 10848-
10853.


CLARIFICATION OF CONTRIBUTIONS

I have been fortunate to have the help of many people in the course of my dissertation work. All work except that described below and cited in the text was exclusively my own.

Chapter Two

The IASPSAD sample (used throughout) was collected by Ken Kendler and Carol Prescott. Genotypic and phenotypic data for replication purposes (also used throughout) were kindly provided by COGA via Danielle Dick. Fazil Aliev assisted in extracting necessary phenotypic information from COGA for craving-related behaviors and AD. Postmortem brain samples were obtained from the NSW TRC at the University of Sydney by Ken Kendler, Gursharan Kalsi and Vladimir Vladimirov. DNA was extracted by Omari McMichael and a GWAS performed by the Vanderbilt core facility. NSW TRC gene expression data was provided by Dayne Mayfield. Statistical analysis of gene expression was performed in conjunction with Vladimir Vladimirov. The gene-set list for pathway analyses was developed with the help of Jill Bettinger and Mike Grotewiel.

Chapter Three

The pooled GWAS in the IASPSAD was performed by TGEN and analyzed by Brien Riley and Po-Hsiu Kuo. Independent genotyping was completed with the help of Nishant Parikh and Jeff Alexander. Fazil Aliev assisted in extracting necessary phenotypic information from COGA for AD. Additional genotypic and phenotypic
information for replication was obtained from the MGS project via Ken Kendler. Factor score phenotypes in the MGS sample were provided by Gursharan Kalsi and Steve Aggen. The postmortem brain samples and analyses are as described above for Chapter Two. The eQTL study was performed by Mike Miles and Andrew van der Vaart. Analysis of ENCODE data was done by Aaron Wolen.

Chapters Four and Five

IASPSAD case samples were quantified and run on gels with help from Laura Hack and Brandon Wormley. A GWAS was performed at the Vanderbilt Core Facility. GWAS data was cleaned by Laura Hack and Todd Webb and genotypes called by Laura Hack. The dataset was imputed by Tim Bigdeli.

Chapter Six

Fazil Aliev assisted with extraction of MD and neuroticism phenotypes from the COGA sample. Factor analysis and invariance modeling code was written in conjunction with Steve Aggen and analyzed with the help of Darlene Kertes.
Amy Elizabeth Adkins was born on June 13, 1983, in Henrico County, Virginia, and is an American citizen. She graduated from Manchester High School, Chesterfield, Virginia in 2001. She received her Bachelor of Science in Biology from Virginia Polytechnic Institute and State University, Blacksburg, Virginia in 2005.