2012

Examining Alcohol Dependence and Its Correlates From A Genetically Informative Perspective

Laura Hack
Virginia Commonwealth University

EXAMINING ALCOHOL DEPENDENCE AND ITS CORRELATES FROM A GENETICALLY INFORMATIVE PERSPECTIVE

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

by

Laura Michele Hack
Bachelor of Science, The College of William & Mary, 2006

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

Kenneth S. Kendler, M.D.
Distinguished Professor, Departments of Psychiatry and Human and Molecular Genetics

Virginia Commonwealth University
Richmond, Virginia
September, 2012

Acknowledgement

First, I'd like to express my sincere appreciation to my advisors, Drs. Brien Riley and Kenneth Kendler, for the enormous amount of time that they dedicated to mentoring me, guiding me through analytic challenges, and fostering my passion for research. I'd also like to acknowledge the other members of my committee, Drs. Danielle Dick, Jill Bettinger, and Tim York, as well as Drs. Todd Webb and Hermine Maes, for their time, guidance, and support. I want to recognize the research associates and post-docs at VIPBG, including Steve Aggen, Fazil Aliev, Jocilyn Dellava Bergin, Tim Bigdeli, Alexis Edwards, Chuck Gardner, John Myers, Elizabeth Prom-Wormley, and Brad Verhulst, for their countless hours aiding me with statistical analyses. Thank you to past and present members of the Riley lab, including Amy Adkins, Dave Brohawn, Rachel Elves, Gursharan Kalsi, Erik Loken, Cathy Siu, Dawn Thiselton, and Brandon Wormley for their personal and professional support and the good times! Additionally, I couldn't have made it through my doctoral training without a great deal of assistance from the administrative and IT support staff. Thank you to all of my VIPBG and non-VIPBG friends for your wonderful friendship. I'd like to recognize Dr. Gordon Archer for the opportunity to do my MD-PhD training at VCU and for his guidance over the years. Thank you to Dr. John Griffin, my undergraduate research mentor, for providing me with my first research experience and inspiring me to continue to pursue research. Finally, I'd like to express my profound gratitude to my wonderful family and boyfriend whose endless support, encouragement, and love have been integral to my success.

Table of Contents

List of Tables

List of Figures

Abbreviations

A.....additive genetic variance
AA.....alcohol abuse
AAsx ..... alcohol abuse symptoms
AC..... alcohol consumption
ABP.....alcoholic beverage preference
AD.....alcohol dependence
ADHD ..... attention deficit hyperactivity disorder
ADHDsx .....attention deficit hyperactivity disorder symptoms
ADsx ..... alcohol dependence symptoms
AIC..... Akaike Information Criterion
ARP..... alcohol related phenotype
ASPD ..... antisocial personality disorder
ASPDsx.....antisocial personality disorder symptoms
AUD..... alcohol use disorder
β .....regression coefficient
BC ..... BEAGLECALL
BP..... base pair
C.....common environmental variance
CI.....confidence interval
CD.....conduct disorder
CTD..... classical twin design
COGA .....Collaborative Study on the Genetics of Alcoholism
D..... non-additive genetic variance
DA.....dopamine
df .....degrees of freedom
DD..... drug dependence
DDsx .....drug dependence symptoms
DSM..... Diagnostic and Statistical Manual of Mental Disorders
DZ .....dizygotic
E .....unique environmental component

ETFD..... extended twin family design

F ..... vertical cultural transmission or F-statistic

FDR..... false discovery rate

FNE.....Fear of Negative Evaluation Scale

FT12.....FinnTwin12

GAD..... generalized anxiety disorder

G x E ..... gene-environment interaction

GWAS..... genome-wide association study

IASPSAD..... Irish Affected Sib Pair Study of Alcohol Dependence

ISENS .....initial sensitivity

λ.....genomic inflation factor

LD ..... linkage disequilibrium

MAF..... minor allele frequency

MAX24 ..... maximum drinks in 24 hours

MD ..... major depressive disorder

MZ.....monozygotic

NS ..... novelty seeking

ONSET..... age of onset of alcohol dependence

OR.....odds ratio

PCA..... principal components analysis

QC..... quanlity control

QQ.....quantile quantile

r_a.....genetic correlation

r_c..... common environmental correlation

r_e..... unique environmental correlation

r_{GE} ..... gene-environment correlation

RAPI .....Rutgers Alcohol Problem Index

SAGE ..... Study of Addiction: Genetics and Environment

SSAGA .....Semi-Structured Assessment for Genetics of Alcoholism

SCID ..... Structured Diagnostic Interview for DSM Disorders

SEM .....structural equation modeling

SES ..... socioeconomic status
SD ..... standard deviation
SNP ..... single nucleotide polymorphism
T ..... special twin environmental variance
TOLMX ..... tolerance/maximum drinking
tSNP ..... tagging single nucleotide polymorphism
VA30K ..... Virginia 30,000
VATSPSUD ..... Virginia Adult Twin Study of Psychiatric and Substance Use Disorders
VEGAS ..... Versatile Gene-based Association Study
WDSFS ..... withdrawal severity factor score

Clarification of Contributions

I greatly appreciate the help of many individuals whose efforts were essential to the generation of this thesis. Their contributions are outlined below.

Chapter 2
Alexis Edwards guided me through the twin analyses and aided in the interpretation of the results. Kenneth Kendler and Elizabeth Prom-Wormley provided insight into the results.

Chapter 3
Hermine Maes, John Myers, and Chuck Gardner guided me through the logistic regression and twin modeling. Brad Verhulst and Jocilyn Dellava Bergin aided with the twin analyses. Hermine Maes and Kenneth Kendler provided insight into the results.

Chapter 4
Kenneth Kendler, Danielle Dick, and Brien Riley helped determine which phenotypes to analyze. Fazil Aliev guided me through the association analyses.

Chapter 5
Amy Adkins, Brandon Wormley, and Erik Loken assisted with the pre-genotyping QC process. Brien Riley, Todd Webb, and Tim Bigdeli guided me through the post-genotyping QC process. Tim Bigdeli was instrumental in the imputation process. Brien Riley, Todd Webb, Kenneth Kendler, and Danielle Dick aided with the phenotypic analyses.

General Abstract

EXAMINING ALCOHOL DEPENDENCE AND ITS CORRELATES FROM A GENETICALLY INFORMATIVE PERSPECTIVE

By Laura Michele Hack, B.S.

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

Advisors:

Brien P. Riley, Ph.D.

Associate Professor, Departments of Psychiatry and Human and Molecular Genetics

Kenneth S. Kendler, M.D.

Distinguished Professor, Departments of Psychiatry and Human and Molecular Genetics

Alcohol dependence (AD) is a serious and common public health problem that contributes to great societal, medical, and legal costs. It has taken work from multiple disciplines, including developmental psychology, genetic epidemiology, and molecular genetics, to achieve our current understanding of environmental and genetic risk factors for AD as well as its variable developmental trajectories. Nevertheless, there is still much to be learned in order to improve treatment outcomes (1). One approach to augmenting our understanding of this disorder is through genetically informative study designs that either examine risk in aggregate or assess specific susceptibility variants. In this dissertation, we utilize both study designs and provide support for the idea that they are both important and useful approaches to continue to pursue.

CHAPTER 1: General Introduction

Significance

Alcohol misuse contributes to 3.8% of global deaths, 40% of traffic accidents, and $234 billion/year in expenses due to medical/legal costs and lost productivity in the US (2,3). Chronic, heavy ethanol consumption is known to cause liver cirrhosis, cardiomyopathy, and pancreatitis in addition to several other medical and psychiatric conditions (4). Furthermore, causal relationships as defined by standard epidemiological criteria (i.e. consistent association and presence of a plausible biological process) have been supported between heavy drinking and a variety of other health conditions, including heart disease, stroke, neuropsychiatric disorders, and several types of cancers (4). Alcohol use disorders (AUDs) are common, relapsing, under-treated, and can have a protracted course. The 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) found the prevalence of lifetime alcohol abuse (AA) and alcohol dependence (AD), respectively, as defined by the Diagnostic and Statistical Manual of Mental Disorders, edition 4 (DSM-IV), to be 17.8% and 12.5%, while the average age of onset for these disorders was 22.5 and 21.9, respectively (5,6). The NESARC also showed that the mean duration of AUDs is 4 years and that only 7.0% and 24.1% of those with lifetime AA and AD, respectively, ever receive treatment (5). Furthermore, for those who are treated, only ~25% of cases remain continuously abstinent one year following treatment (1). Clearly, a better understanding of these serious personal and public health problems is needed to develop more effective prevention and intervention strategies.

Etiology

There is much to be learned about the etiology of AD, which is the more severe AUD, but twin, developmental, and molecular studies over the last few decades have provided some understanding. Twin studies have estimated the genetic contribution to phenotypic variation, or heritability, for AD to be between 50% and 60% (7,8), while the heritability of alcohol consumption (AC) measures and problems ranges from 40%-70%, depending on the phenotype and the study considered (9-11). The remaining phenotypic variation is accounted for by common (i.e. shared by the family) and unique (i.e. limited to one member) environmental factors. A variety of environmental and psychological factors have been identified as correlates of AD, including peer culture, family-related variables, religion, education level, and other psychiatric conditions (5,12). The relationships between AD and its correlates are complex and there is support for direct causal effects as well as shared genetic and environmental liabilities. These two types of etiologic relationships are not mutually exclusive. For example, there is evidence that childhood parental loss and major depression (studied in women) are direct risk factors for AD and share genetic liability with AD (13-15). In this dissertation, we will explore the relationship between AD/ related phenotypes and several of these environmental, psychological, and genetic variables, including education level, alcoholic beverage preference, and DNA polymorphisms.

Phenotypes

The DSM-IV defines AD as follows:

A maladaptive pattern of alcohol use, leading to clinically significant impairment or distress, as manifested by three or more of the following seven criteria, occurring at any time in the same 12-month period:

1. Tolerance, as defined by either of the following:
    - A need for markedly increased amounts of alcohol to achieve intoxication or desired effect.
    - Markedly diminished effect with continued use of the same amount of alcohol.
2. Withdrawal, as defined by either of the following:
    - The characteristic withdrawal syndrome for alcohol (refer to DSM-IV for further details).
    - Alcohol is taken to relieve or avoid withdrawal symptoms.
3. Alcohol is often taken in larger amounts or over a longer period than was intended.
4. There is a persistent desire or there are unsuccessful efforts to cut down or control alcohol use.
5. A great deal of time is spent in activities necessary to obtain alcohol, use alcohol or recover from its effects.
6. Important social, occupational, or recreational activities are given up or reduced because of alcohol use.
7. Alcohol use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the alcohol (e.g., continued drinking despite recognition that an ulcer was made worse by alcohol consumption).

It is clear from this definition that AD is a clinically heterogeneous phenotype that includes a combination of physiological, cognitive, and psychosocial symptoms. Two individuals can be affected by the same disorder but endorse an entirely different set of symptoms. Furthermore, twin research has found that AD symptoms (16) as well as consumption measures and problems (10) comprise multiple dimensions of genetic risk, meaning that different sets of genes contribute to the risk for different aspects of the disorder. Therefore, while the diagnosis of AD continues to be used in gene identification efforts because it offers the advantage of a standardized measure across sites and studies (17), the diversity of symptom and problem combinations in a cohort of subjects can hinder the discovery of genes that contribute to particular sets of symptoms or types of problems**.** One strategy to overcome this issue is to use stable, heritable, disease-associated quantitative traits as outcomes in association studies (18). If these traits have particular characteristics, including good psychometric properties (e.g. validity and high test-retest reliability), they may be referred to as endophenotypes or intermediate

phenotypes (19). Endophenotypes offer a number of advantages over dichotomous diagnoses. First, they can increase power in genetic association studies because they harbor greater variation than a dichotomous diagnosis (20). Additionally, since some investigators hypothesize that these traits are influenced by a smaller number of genetic loci than diagnoses, each locus would theoretically have a greater effect size, meaning it would account for a greater proportion of the variance in the trait (21). Several candidate gene association studies that have used intermediate phenotypes such as alcohol craving, initial sensitivity, and maximum number of drinks in 24 hours, have produced signals in genes that did not reach significance in analyses of AD, possibly due to greater effect sizes of these endophenotypes (22-25). Second, intermediate phenotypes are more useful in community samples, where individuals with a diagnosis represent a minority of the subjects (20). Finally, some endophenotypes may be tested in model organisms, whereas AD cannot.

Another approach to gene discovery that has been explored in the literature is consideration of other psychiatric disorders and behavioral traits that are often comorbid with alcohol dependence. These phenotypes are broadly classified into two groups: (1) internalizing, which includes neuroticism, depression, and anxiety, and (2) externalizing, which includes drug use disorders, conduct disorder, antisocial personality disorder, and a disinhibited personality (26,27). Twin studies suggest that there is strong genetic overlap between externalizing disorders and AD, often considered part of the externalizing spectrum, and small to modest genetic overlap between internalizing disorders and AD (5,15,28-30). Previous work has demonstrated the utility of considering broader phenotypes in gene discovery efforts. For example, Dick and colleagues found that *CHRM2* was more strongly associated with an externalizing component score than

individual phenotypes that comprised the score, suggesting that this gene contributes to risk via a broader externalizing pathway rather than through a mechanism that is specific to AD (31).

Study Design

Research in the fields of epidemiology, psychiatric genetics, and developmental psychology has suggested that AD and alcohol related phenotypes (ARPs) involve (1) the main effects of genes and the environment; (2) interaction between these components; (3) the variable effect of these components across time within individuals, and (4) variable developmental trajectories across individuals. In order to gain a more complete understanding of these highly complex phenotypes, researchers must use a variety of study designs involving humans and model organisms, rich phenotypic, genotypic, and environmental information, and measures across time. In this dissertation, we focus on two approaches: genetic epidemiology (specifically twin studies) and gene discovery studies. The former allows us to gain a sense of aggregate genetic and environmental risk, while the latter enables identification of specific susceptibility loci (32). While some argue that twin studies are no longer needed in an age genomics, they remain the only way to understand risk in aggregate (33) and can be used to inform association studies in a number of ways. First, researchers can use them to identify phenotypes/ endophenotypes that are genetically correlated with diagnoses but are more easily collected and widely available. For example, Kendler et al. showed that AC measures assessed in a population sample during the heaviest year of drinking have high genetic correlations with AD, suggesting that these measures are good proxies for AD in gene discovery efforts (9). Second, twin studies can be used to construct alternative phenotypes that are potentially more genetically homogeneous and reduce the number of phenotypes tested. In an unpublished study by Dick and

colleagues, investigators showed that two latent genetic factors accounted for the genetic variation in seven measures of alcohol use/problems, and they constructed genetic factor scores from the factor loadings for use in an association study. They found several associations between the genetic factor scores and SNPs within *DRD2/ANKK1* but none with the original phenotypes, suggesting that use of the scores may have indeed reduced genetic heterogeneity. Third, if twin models suggest that measured environmental variables moderate a trait's effect on latent genetic influences (latent G x E), this is good reason to incorporate these environmental variables into gene discovery studies, as at least a good portion of the genes influencing that trait will be moderated by the environmental variable (34). Unfortunately, candidate gene studies (candidate G x E) of psychiatric disorders that have attempted to incorporate measured environmental variables have suffered from low power and many false positive findings (32,35). However, testing moderation of a polygenic score of GWAS findings appears more promising, as this approach does not rely on the selection of the proper gene/environment combination.

While study paradigms will be discussed more extensively in subsequent chapters, they are briefly introduced here. The basic twin design compares the phenotypic similarity of monozygotic and dizygotic twins to determine the amount of phenotypic variation that can be attributed to additive genetic as well as common and unique environmental factors. Many extensions to twin studies exist that allow researchers to ask more complex questions about sources of variation, such as (1) Are the magnitudes of the sources of phenotypic variation different between the sexes? (2) What other types of variation (e.g. a special twin environment) contribute to the phenotype of interest? (3) Does the effect of genes alter depending on environment or, equivalently, does the effect of the environment change depending on genes? (4) How do the magnitudes of the sources of variation change throughout time? In Chapters 2 and 3,

we focus on answering questions (1)-(3) for several ARPs (consumption measures and AUD symptoms) and their correlates (education level, alcoholic beverage preference, and other psychiatric disorders).

When moving from evaluation of aggregate genetic risk to specific variants, there are several options for study designs, including hypothesis-driven candidate gene approaches as well as agnostic genome-wide association studies (GWAS). In Chapters 4 and 5, we use both approaches to help us identify risk variants for AD and related phenotypes. Although costs are now fairly reasonable for genome-wide genotyping arrays (~$300/individual), researchers continue to pursue both approaches because each has its own advantages and disadvantages. Candidate gene studies allow investigators to select particular variants of interest instead relying on chip manufacturer's decisions, they do not have the same multiple testing burden as GWAS, and they are less prone to genotyping error. GWAS are advantageous in that researchers can avoid selection bias, identify candidates that were not previously selected for biological or other relevant reasons, more easily control for population stratification, and use their money more efficiently. Both approaches generally interrogate common variants, or those with minor allele frequencies (MAF) of greater than 5% for candidate gene and smaller genome-wide arrays, and greater than 1% for larger GWAS arrays (36). For GWAS, it is estimated that between 500,000 and 1 million markers are needed to capture common variation in European populations due to the correlated nature of the genome, known as linkage disequilibrium, or LD. Therefore, most commercial genotyping arrays contain SNPs within this range.

GWAS

GWAS for non-psychiatric disorders have generally been successful at identifying replicable variants that were not previously detected using biological or other relevant criteria. For example, prior to the GWAS era, there were only a handful of independent loci robustly associated with Type II Diabetes (T2D) and related metabolic traits, but GWAS have increased this number to 75 (37). Psychiatric disorders have lagged behind other complex diseases for a number of reasons. First, psychiatric disorders are non-ideal phenotypes for genetic inquiry. While many diseases have physiological or histological metrics to categorize cases (e.g. fasting plasma glucose >126 mg/dl in T2D), the diagnostic criteria for psychiatric disorders are interview-based, and thus, more subjective both on the part of the interviewer and the respondent. Furthermore, the DSM criteria were developed primarily for clinical purposes and are not well suited to gene identification efforts, since twin studies have found that they show multiple dimensions of genetic liability (16,38). Second, many GWAS of psychiatric disorders have inadequate sample sizes for the effect sizes expected (OR=1.1-1.5). However, as the sample sizes in GWAS of psychiatric disorders have increased, so have the number of loci discovered. Recent, well-powered GWAS of schizophrenia report multiple SNPs reaching the threshold for genome wide significance, and several discovery as well as replication samples have implicated the same genes, particularly the MHC region on chromosome 6 (39). GWAS of AD and ARPs have not reached the sample sizes, either individually or collectively, that psychiatric phenotypes like schizophrenia and smoking have, and this is one reason they may not have been as successful thus far. However, with growing numbers of GWAS of AD and related quantitative traits and the work of the alcohol GWAS consortium, which includes over 20,000 subjects, there is optimism that alcohol dependence will begin to experience similar success. Finally, some

GWAS of high prevalence psychiatric disorders have used unscreened controls, which reduce power.

We addressed each of these issues in our association study detailed in Chapter 5. For phenotypes, we examined not only AD, but also two quantitative traits: AD symptom count and an externalizing composite phenotype. While our sample is moderately sized, as most individual samples are, it is part of the alcohol GWAS consortium, which is better powered to identify effects at the lower end of the odds ratio spectrum. For financial reasons, we used a lightly screened sample of blood donor as controls for a high prevalence disorder. However, there is evidence that non-remunerating blood donors have lower alcohol intake than the general population (40), and thus likely a lower prevalence of AD. Additionally, we accounted for the fact that the controls were only lightly screened by coding them as unknown and inputting an estimated prevalence of AD for blood donors in our binary analysis.

Correction for Multiple Testing

In many candidate gene and all GWA studies, the issue of multiple testing must be addressed to limit Type I error, or the number of false positives. In Bonferroni correction, the α level is adjusted through division of the target α level (usually 0.05) by the number of tests conducted. Because Bonferroni correction assumes all tests are independent and there are typically many correlated markers on genome-wide genotyping arrays, efforts have been made to estimate the effective number of independent tests in the genomes of several ancestry groups. For European populations, these estimates are 388,751 SNPs for the Affymetrix 6.0 genotyping array and ~1.6 million SNPs for imputation to 1000 Genomes, corresponding to *p*-values of 1.08×10^{-7} and 3.06×10^{-8}, respectively (41). While many candidate gene and GWA studies use a

Bonferroni correction, some argue that it is too stringent and will lead to many missed findings (42). Bonferroni correction controls the family-wise error rate, or the probability of obtaining at least one false positive among *all* tests, while controlling the false discovery rate (FDR) limits the expected proportion of false positives among *significant* tests. Many authors suggest that in GWAS studies, it is desirable to allow for some false discoveries in order to maximize the number of true effects observed. In Chapter 5, we utilize the FDR method of Storey in which *p*-values are ranked in descending order and *q*-values are estimated according to a formula that is discussed later in more detail (43). For a particular test, if the *q*-value is less than a pre-determined threshold of 0.2, for example, then 20% of the tests with *q*-values $\leq$ 0.2 are false discoveries.

Prior Evidence from Association Studies of AD

Hundreds of genes have been implicated in candidate gene and GWA studies of AD and ARPs. Here we review genes that show evidence across multiple studies or are suggestive/ genome-wide significant. The most robust evidence has come from the alcohol metabolizing genes. Additionally, because alcohol has widespread and diverse effects throughout the brain, a large number of genes encoding proteins in neurotransmitter systems have been tested and show good evidence of involvement in AD.

ADH & ALDH

Alcohol is mainly metabolized in the liver by alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH). ADH converts alcohol into acetaldehyde, a toxic metabolite that causes facial flushing and nausea when it accumulates, while ALDH oxidizes acetaldehyde to the harmless substance acetic acid. The *ADH* gene family consists of seven members encoded

on chromosome 4q21-23. All three members of class I (*ADH1A,1B, 1C*) and the single member of class II (*ADH3*) encode enzymes that function in ethanol metabolism, while *ADH4-AHD6* metabolize longer-chain alcohols (44). Two class I functional variants, *ADH1B*2* (His47) and *ADH1B*3* (Cys369), are common in East Asian/certain Jewish populations and Africans/Native Americans, respectively, and have high enzymatic activity (45). These variants have been associated with lower risk for AD in both candidate gene and GWA studies in Asians and Europeans (46,47). It is hypothesized that these variants are protective against alcohol dependence because they increase the rate at which ethanol is converted to a toxin, therefore, decreasing the likelihood that an individual would want to ingest ethanol (48). However, it has been shown that with normal functioning of the next step in the metabolism process, this variant does not lead to appreciable accumulation of acetaldehyde; thus, the mechanism of protection is still unclear (49). The remaining classes are less well studied, but *ADH4* variants have been implicated in AD in European and African Americans (50,51). Nineteen potentially functional genes and 3 pseudogenes in the ALDH superfamily have been identified in humans, but only two (*ALDH1* and *ALDH2*) encode isozymes that are thought to important in acetaldehyde metabolism. *ALDH1* is expressed in the cytosol of liver cells while *ALDH2* is expressed in the mitochondria. Individuals heterozygous for the functional *ALDH2*2* (Lys487) allele, which is common in Asian populations, have 30%-50% enzymatic activity while homozygotes have completed ablated activity. This polymorphism has been associated with reduced risk of AD and dramatic increases of acetaldehyde in the blood (44,49,52). Therefore, it is thought that this variant is protective due to the large and prolonged accumulation of the unpleasant substance, acetaldehyde. Additionally, a recent GWAS in a Japanese population identified a highly

significant association between a rare nonsynonomous coding polymorphism in *ALDH2* (Lys504) and alcohol consumption (53).

The fact that variation in the metabolizing genes is *protective* suggests a different model of disease for AD than the classic model in which mutation → dysfunction. Genes that efficiently metabolize ethanol were advantageous in an environment in which ethanol was only present in sources like fermenting fruit. In our modern environment where ethanol is much more widely available, these genes have become deleterious. This is a similar to the disease model of T2D because in ancient environments where sources of energy were scare, genes that help to efficiently store fat were advantageous. However, in the modern environment of excessive food sources, these genes contribute to the development of T2D. Thus, an additional reason to believe that GWAS of sufficient size will begin to produce replicated risk loci for AD is that GWAS of TD2 have been very successful.

GABAergic System

γ-aminobutyric-acid (GABA) is the brain's main inhibitory neurotransmitter and is thought to mediate several responses to alcohol, including sedation, motor incoordination, tolerance, and dependence. Ethanol may cause these effects by altering the level of GABA and the firing rate of its neurons as well as changing the expression levels of $GABA_A$ receptor subunits. $GABA_A$ receptors are composed of five out of a possible 19 subunits, including $\alpha_{1–6}$, $\beta_{1–3}$, $\gamma_{1–3}$, δ, ε, θ, π, and $\rho_{1–3}$. The most common configuration is two α, two β, and one γ subunit. At anesthetic concentrations, ethanol binds to the α_1 subunit and causes conformational changes within the receptor, but this effect has not been demonstrated at physiological concentrations (54). Genes encoding the α, β, and γ, subunits are *GABRA1-6, GABRB1-3*, and *GABRG1-3*,

respectively. There is a cluster of GABA receptor genes on chromosome 4p, including *GABRA2, GABRA4, GABRB1,* and *GABRG1,* and this region has been implicated in several linkage studies of AD and β EEG power, an endophenotype for AD (54). Multiple studies have also found associations with SNPs and haplotypes in *GABRA2* and AD (54). There is less evidence of association with the chromosome 5q (*GABRA1*, *GABRA6, GABRB2*, *GABRG2)* and 15q (*GABRA5, GABRB3, GABRG3*) clusters, but *GABRA1* was associated with AD and measures of drinking severity in the Collaborative Study on the Genetics of Alcoholism (COGA) (55).

Dopaminergic System

Multiple types of animal studies, including microinjection, microdialysis, and lesioning, have implicated dopamine in mediating the rewarding effects of ethanol (56). A meta-analysis found the commonly studied TaqIA restriction fragment length polymorphism (rs1800497) to be associated with AD at a modest OR of 1.31 (57) . While this SNP is located within the coding region of the neighboring *ANKK1* gene (58), it may be tagging a polymorphism in *DRD2*, which contains several SNPs in LD ($r^2 \approx 0.7$) with rs1800497. Dick et al. found weak associations between variants in *DRD2* and AD as well as stronger associations with neighboring genes *ANKK1*, *TTC2,* and *NRCAM1* (59). Rigorous studies of *DRD1* and AD have typically found positive associations with modest effect sizes (60,61), while reports examining this phenotype and *DRD3* have been negative (57). A review supports an association between the 48-bp variable number of tandem repeats (VNTR) in *DRD4* and an intermediate phenotype termed urge for addictive substances, which refers to craving for substances of abuse (62). Our report in Chapter 4 was the first to show associations between *DRD5* and an ARP, an alcohol withdrawal factor score.

Other Top and/or Replicated Genes from GWAS

To our knowledge, there have been six GWAS of AD (46,63-67) and four of AC measures and symptoms (68-71). In addition to a SNP located between *ADH1B* and *ADH1C* (46), two intergenic SNPs in LD reached genome-wide significance in GWAS of AD in a German sample (63). These SNP are near the gene *PECR*, which is a key enzyme in fatty acid metabolism. One SNP in *AUTS2*, or autism susceptibility candidate 2 gene, was found to be genome-wide significant in a GWAS of alcohol consumption in Europeans (70). The function of this gene is unknown, but investigation by Schumann and colleagues in Drosophila showed evidence that *AUTS2* is involved in regulating response to ethanol (70). In a GWAS of Korean men, several SNPs greatly surpassed the level of genome-wide significance for alcohol consumption, including those in the following genes: *C12orf51, CCDC63, MYL2, OAS3, CUX2,* and *RPH3A* (71). Some of these significant SNPs were in high LD with SNPs in *ALDH2.*

Genes containing SNPs in the genome-wide suggestive range of 5×10^{-6} for GWAS of AD include: *PPP2R2B, CTBP2, KRT3, TJP1, PKNOX2, CC2D2B, SH3BP5, GRM5,* and *BBX* (72). Thus far, there have been some examples of replication of genes from one GWAS of AD to the another, including *ERAP1, CPE , DNASE2B, SLC10A2, ARL6IP5, ID4, GATA4 , SYNE1,* and *ADCY3* (64,65). For genes that have known functions, they broadly fall into the following categories: cell growth (*PPP2R2B*), transcription (*CTBP2, PKNOX2, BBX, ID4, GATA4*), keratin formation (*KRT3*), signaling (*TJP1, SH3BP5, ADCY3, PRKCA*), reception (*GRM5*), enzymatic activity (*CPE*), and transport (*ARL6IP5, SLC10A2*).

Overview of Current Research

The aim of this dissertation is to expand our understanding of the etiology of AUDs and related traits using genetically informative study designs. We take advantage of both the global perspective provided by genetic epidemiology as well as the more focused contribution of association studies. Part I (Chapters 2 and 3) uses the twin approach to explore the relationship between ARPs and education level as well as the correlates and genetic architecture of alcoholic beverage preference. In Part II (Chapters 4 & 5), we test for loci that contribute to risk for AD, AC measures, and phenotypes that are part of the externalizing spectrum. Some of the findings we present are currently being pursued using study designs that were not utilized in this dissertation (i.e. targeted and whole genome sequencing, and use of population samples and model organisms) to further our understanding of these complex and daunting phenotypes.

CHAPTER 2: Exploring the Association between Education Level & Alcohol Related Phenotypes Using a Genetically Informative Design

Abstract

Background: An extensive literature consisting of both cross-sectional and longitudinal studies has examined the relationship between educational attainment and a variety of alcohol use measures, finding both positive and negative statistical relationships. However, few studies have done so from a genetically informative perspective. A recent Finnish twin report assessed the extent to which these correlations could be explained by genetic and environmental influences on both phenotypes using two alcohol problem measures (73). They estimated modest, negative genetic correlations between educational attainment and both alcohol-related traits, but no environmental correlations. Additionally, they showed that education level significantly moderated genetic influences on alcohol problems such that higher education was associated with greater heritability of alcohol measures.

Methods: We undertook a similar study in a Virginia-born sample of twins using alcohol dependence symptoms (ADsx) and maximum quantity in a 24 hour period as our alcohol measures. Descriptive statistics were generated in SAS and twin analyses were conducted in classic Mx.

Results: We found modest negative phenotypic correlations in males and non-significant negative phenotypic correlations in females between educational qualifications and alcohol measures. We also estimated a significant negative genetic correlation between educational attainment and maximum quantity, but not ADsx, in males. Furthermore, unstandardized

parameters estimates showed that unique environmental variance decreased with increasing level of education, whereas genetic variance decreased slightly or remained stable in both sexes. To account for the change in unique environmental variance, when standardized, the heritabilities of ADsx and maximum quantity increased with increasing levels of education, reflecting the results in the Finnish twins.

Conclusions: Our results have implications for the long standing debate about the relationship between socioeconomic status, a strong correlate of education, and mental health. Moreover, our work underscores the importance of considering unstandardized parameter estimates when interpreting the results of gene-environment interaction studies.

Introduction

Over the past two decades, several longitudinal studies in the US and abroad have found significant associations between (a) level of educational attainment and subsequent development of drinking patterns, alcohol use disorders, and alcohol related problems (74-76) as well as (b) alcohol misuse and subsequent educational attainment (77,78). However, the relationship is complex and not entirely consistent between studies; correlations vary widely in both strength and direction, depending on the sex, age, and country of origin of the participants as well as the outcome studied (e.g. drink frequency, drink quantity, alcohol abuse, etc.) (76). Generally, these associations remain after controlling for correlates of both educational level and alcohol measures; therefore, several causal explanations have been proposed. If the correlation is negative and the temporal ordering is such that the educational attainment occurs before the development of alcohol problems, one causal explanation is that lack of education, which may equate to unfulfilled social expectations, is stressful and individuals cope by developing problematic drinking that eventually leads to alcohol problems (74). Assuming the same

chronological order, positive correlations may be explained by the fact that more educated individuals generally have more disposable income with which to purchase alcohol and may be exposed to work-related situations in which drinking is acceptable and encouraged (e.g. happy hours) (76). Finally, if the correlation is negative and the alcohol misuse occurs before the educational attainment, then some causal explanations include direct ethanol toxicity on cerebral functioning or the reduction/cessation of social bonding important for encouragement of the continuation of education (77).

Another reason for the significant correlations between educational outcomes and alcohol measures may be that at least some of the same genes contribute to variation in both phenotypes. Individually, both phenotypes have been found to be moderately to strongly heritable. Heritability, or proportion of phenotypic variation accounted for by genes, varies based on which alcohol related phenotype (ARP) one is considering, but is generally on the order of 40-60% (8). The heritability of educational attainment has been estimated at 31%-82% (79-81). One will note the fairly wide range for these estimates, particularly for education, depending on the age, sex, and cohort being studied.

In a Finnish sample of twins, Latvala and colleagues examined the non-mutually exclusive hypotheses that (1) education and alcohol problems are genetically correlated (i.e. share predisposing genes) and (2) the importance of genetic and environmental influences on variation in alcohol problems varies depending on level of educational attainment (73). The authors used two measures of alcohol problems: (a) a modified version of the Rutgers Alcohol Problem Index (RAPI), a self-report measure that assesses alcohol problems in the previous year and (b) a self-reported estimate of the maximum number of alcoholic beverages consumed within a 24 hour period throughout one's lifetime. Latvala and colleagues first estimated modest

but significant *negative* phenotypic correlations between their alcohol measures and education level in both sexes. Next, they assessed whether these phenotypic correlations could be at least partially explained by overlapping genetic factors. In both sexes, they estimated modest significant *negative* genetic correlations between education and both RAPI (females= -0.22; males= -0.12) and maximum drinks (females= -0.17; males= -0.19). These results suggest that, indeed, there is a common set of genes that increase the risk of both low educational attainment and alcohol problems. Next, the authors conducted a set of univariate twin analyses in which they assessed whether educational attainment moderated genetic (and environmental) influences on their outcomes, called gene-environment (and environment-environment) interaction (G x E, E x E). In other words, they asked if the influence of genetic predispositions for alcohol problems varied as a function of education level (or equivalently, the importance of educational level on the risk for alcohol problems varied as a function of genotype) (82). Several previous studies have found that the heritabilities of substance related traits decrease when individuals are involved in marriage-like relationships (83), take part in prosocial activities (84), live in rural environments (85,86), have decreased exposure to deviant peers (87,88), and experience a high level of parental monitoring (89,90). In both sexes, Latvala and colleagues found that education level moderated the unique environmental influences on RAPI and maximum drinks such that the unique environment had a greater influence on variation in alcohol problems in less educated individuals (73). Although moderation of genetic effects was not significant, in order to compensate for the relative change in unique environmental variation, heritability of alcohol problems varied such that it was higher in those with greater educational qualifications. We conducted a similar study in a Virginia-born sample of twins to determine if these results would generalize to a cohort with different cultural influences and a wider age range.

Methods

Subjects

We used data collected from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD), a longitudinal study of psychiatric and substance-related disorders in 2 samples of adult twins identified through the population-based Virginia Twin Registry born from 1934-1974. The subjects in the first study are comprised of female-female twin pairs, while the subjects in the second study include both male-male and male-female twins. For further details of recruitment and subject characteristics, see Kendler and Prescott (91). We utilized data for the present report from the fourth wave of the female-female study (FF4) and the second wave of male-male and male-female study (MM/MF2). After receiving a description of the protocol, participants provided written consent for in-person interviews and verbal consent for telephone interviews. The Office of Research Subjects Protection at Virginia Commonwealth University approved this project.

Measures

All measures were collected by clinically trained interviewers. RAPI was not used in the VA twins; however, DSM-IV alcohol-related diagnoses were assessed in this sample (6). We elected to use ADsx because this measure has a greater range (0-7) than abuse symptoms and it is more often utilized in genetically informative study designs. Symptoms counts were collected using an adaptation of the Structured Clinical Interview for DSM-IV Disorders (92). Additionally, we included a measure similar to the maximum drinks variable in the Finnish twins; the difference was that our measure was in reference to the heaviest year of drinking, while the measure in the Finnish study represented a lifetime measure. Maximum quantity was

log transformed for use in our twin models. Educational attainment was assessed by asking the question, "What is the highest grade of school or year of college you completed?" Answer choices ranged continuously from 0 years to 20+ years. We converted education to z-scores (abbreviated as zEduc) for ease of interpretation in our moderation models.

Descriptive Statistics & Model Fitting

Descriptive Statistics. Descriptive statistics were obtained separately for males and females using raw scores in SAS 9.2 (93). To assess whether associations between our alcohol measures and educational attainment existed in our sample of twins, we calculated phenotypic correlations separately by sex using one twin randomly chosen from each twin pair.

Classical Twin Design. The classical twin design (CTD) attempts to estimate the genetic and environmental contributions to twin-pair resemblance (94). By considering the phenotypic correlations between different types of twins in a pair, one can determine the relative importance of three sources of latent variation: additive genes (A or a^2), common environment (C or c^2), and individual-specific environment (E or e^2). A refers to variation resulting from allelic effects across multiple genes; C reflects experiences that both twins have in common and that make them more similar; and E consists of both environmental experiences not shared by twins as well as random measurement error (95). We assume that monozygotic (MZ) twins are correlated at 1 and dizygotic (DZ) twins at 0.5 for their genes; that is, they share 100% and 50% of their genes in common, respectively. Additionally, both types of twins have a correlation of 1 for their shared environment and no correlation for the unique environment. These correlations are illustrated in diagrammatic form in Figure 2.1. Using these assumptions, basic path analysis

provides us with the following equations for the correlations (standardized covariances) between MZ and DZ twin pairs:

$$\begin{aligned} rMz &= A + C \\ rDZ &= 0.5A + C \end{aligned}$$

Solving these equations for A and C allows for their estimation. Because the three standardized sources of variation by definition must sum to 1, E is estimated by subtraction of *rMZ* from 1:

$$\begin{aligned} A &= 2(rMZ - rDZ) \\ C &= rMZ - A \\ E &= 1 - rMZ \end{aligned}$$

If *rMZ* is greater than *rDZ*, genetic factors likely play an important role in the phenotype, while if *rDZ* > ½ *rMZ*, shared environmental factors are probably influential. Finally, if *rDZ* < ½ *rMZ* then non-additive genetic effects (D) likely play a role in the outcome; however, both C and D cannot be estimated simultaneously in a sample of twins only. E is always present because this term includes measurement error.

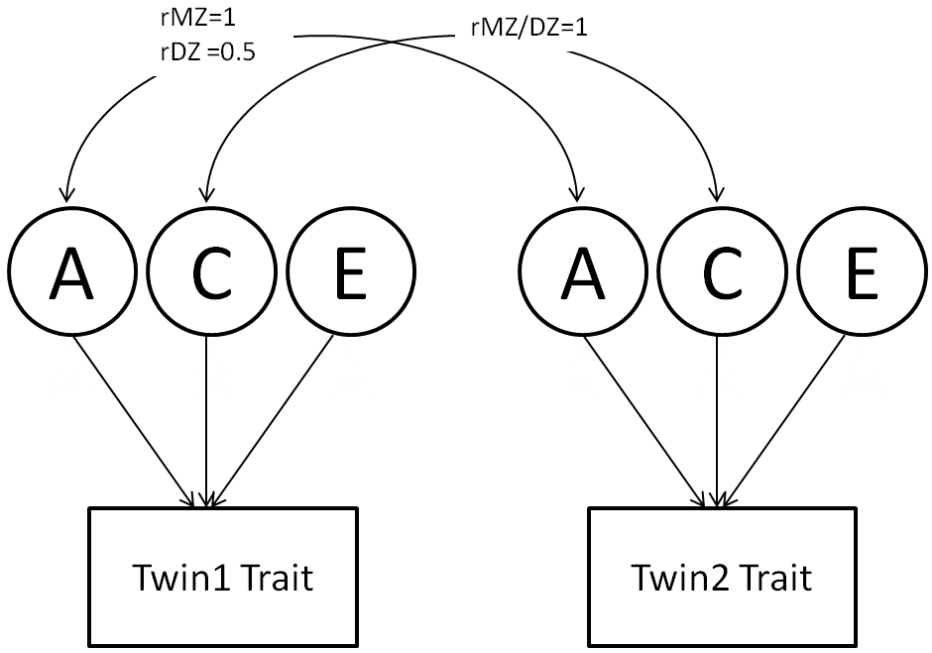

Figure 2.1. Path diagram of univariate model. rMZ, correlation between MZ twins; rDZ, correlation between DZ twins

Structural Equation Modeling. While calculating A, C , and E using these equations is quick, the preferred approach to estimation of these parameters is structural equation modeling (SEM), which, among other advantages, allows for evaluation of the significance of each parameter, inclusion of covariates, and extension to include multiple phenotypes. In essence, a SEM is a linear regression model that involves two types of statistical relationships between variables: correlational and causal. The phenotypic relationships between twins in a pair are correlational, while the influences of A, C, and E on the variation of the phenotype of interest are causal (96). Model fitting proceeds by iteratively improving upon parameter starting values until they have been optimized. There are many approaches to numerical optimization, but we selected full information maximum likelihood because it is most commonly used in analyses with twin data and it provides a convenient way of handling missing data (97,98). Parameters may be successively dropped (i.e. set to zero) to evaluate their significance in the model. The fit of a model to the observed data is summarized by a statistic, -2*log likelihood, which is distributed as a χ^2. If the $\Delta\chi^2$ between two models has a significant *p*-value, the nested model is rejected, and the tested parameter is retained in the model. Model fit may also be assessed by a variety of fit indices, including Akaike's information criterion (AIC), Bayesian information criterion, and root mean square error approximation. We chose to examine fits using the commonly applied ΔAIC, which equals the difference in -2 log likelihood between the models minus the difference in the degrees of freedom multiplied by 2, or $\Delta\chi^2$-Δ2df (degrees of freedom). A lower AIC suggests a better balance between parsimony and explanatory power (99,100), or better model fit.

Cross-Twin Cross-Trait Correlations. Prior to formal twin modeling, we assessed cross-twin cross-trait correlations to gain a sense of whether shared genetic factors contribute to the phenotypic correlations between education and our alcohol-related traits. These correlations

provide an estimate of the association of education level in one member of a twin pair with each alcohol measure in the other. They were calculated using the double-entry method in which each twin was entered as both Twin 1 and Twin 2, so that the variation associated with the arbitrary ordering of twins can be removed. Similar to assessment of correlations for univariate analyses, if the cross-twin cross-trait *rMZ* > *rDZ*, additive genetic factors shared between phenotypes is likely important. If *rDZ* > ½ *rMZ,* shared C probably contributes to the phenotypic correlation.

Model Fitting. Thorough model fitting begins with a saturated model, which allows for testing of the CTD assumptions that means and variances (continuous traits) or thresholds (ordinal traits) are independent of birth order, zygosity, and sex. If these assumptions do not hold true and are not accounted for in the model, the results may be biased (97). When trait data is continuous, as it is in our case, we test the assumptions that there are no differences in the means and variances between Twin 1 and Twin 2 in each pair, MZ and DZ twin pairs, and males and females by comparing the fit of models in which these components are estimated separately to the fit of models in which they are equated. We performed all model fitting using the SEM program Mx (101) using four groups: female and male MZ and DZ twins. We did not include opposite sex twins because they add to the complexity of model fitting. In order to test the assumptions and to obtain good starting values for our bivariate moderation models, we began by fitting the data to univariate models for each trait.

We then investigated whether education moderates the genetic and environmental influences on the ARPs in a series of bivariate moderation models. If gene-environment correlations (r_{GE}) are present between the moderator and the trait being studied, this can inflate the level of moderation. r_{GE} occurs when an environmental variable is not "pure", but rather is itself influenced by genes that also influence the trait under study (102). Given that negative

genetic correlations were found between educational level and alcohol problems in the Finnish study (73), we hypothesized that this would also be the case in our data. We found negative genetic correlations using standard bivariate twin modeling across education and the two ARPs in both sexes; however, only the genetic correlation between maximum quantity and education in males was significant (data not shown). Purcell has developed two models to address G x E in the presence of $r_{GE.}$ In one model, the effect of the moderator is partialled out of the means model, so its effect is only considered on the variance components that are unique to the outcome (i.e. genetic influences on alcohol related traits that do not overlap with genetic influences on educational attainment) (103). In the other option developed by Purcell, r_{GE} is explicitly modeled. If present, this allows one to distinguish between moderation of the genetic and environmental influences that are unique to the trait under study and those that are common to both the moderator and the trait, whereas the former model only allows detection of moderation on unique influences. In the path diagram shown in Figure 2.2 for the genetic component only, the "a_{zEduc}" path represents

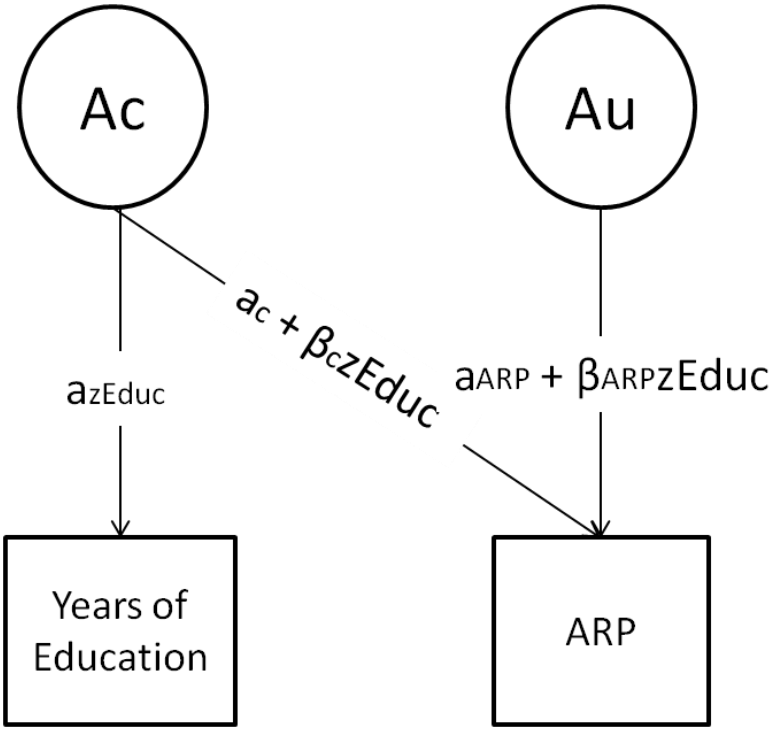

Figure 2.2. Path diagram of moderation model. Genetic component only is shown for Twin 1. The model has the same structure for the common and unique environmental variables as well as for Twin 2.

genetic influences unique to education, "ac" represents genetic influences that are shared between education and ARP, and the "a_{ARP}" path is unique genetic influences on the ARP. Moderation is modeled by the addition of the β_czEduc and β_{ARP}zEduc terms to the path between the traits and the unique path to the ARP. A moderating effect is indicated if any β term is estimated to be significantly different from zero. One advantage of using continuous traits (instead of ordinal ones) in moderation models is that both unstandardized and standardized (forced to sum to 1) parameter estimates can be considered. Unstandardized estimates provide a better sense of how variance is actually changing. For example, it is possible that all three sources of a trait's variation (and, therefore, total variation) change in the same direction, but this would not be evident from a graph of standardized parameter estimates.

Table 2.1. Descriptive statistics for education and alcohol measures

Variable	Females N=3039	Males N=4017
Education Level[a], mean (SD)	14.1 (2.3)	13.5 (2.7)
ADsx[a], N (%)		
0	2424 (79.8)	2232 (55.6)
1	77 (2.5)	220 (5.5)
2	205 (6.8)	469 (11.7)
3	137 (4.5)	380 (9.5)
4	104 (3.4)	301 (7.5)
5	45 (1.5)	213 (5.3)
6	31 (1.0)	132 (3.3)
7	(0.53)	70 (1.7)
Maximum Quantity[a], mean (SD)*	5.6 (4.8)	12.8 (9.6)

Notes: ADsx, alcohol dependence symptoms. [a]Significant difference at $p<0.05$ as assessed by ANOVA for continuous variables and Pearson's χ^2 for categorical variables.

Table 2.2. Pearson correlations between education and alcohol measures

ARP	Females (95% CI) N=3039	Males (95% CI) N=4017
ADsx	-0.04 (-.09,.01)	**-0.13 (-.16, -.09)**
Maximum Quantity	-0.04 (-.09, .01)	**-0.16 (-.20, -.12)**

Notes: ADsx, alcohol dependence symptoms. Significant correlations in bold. Correlations calculated using one randomly chosen twin from each pair. Raw values for all variables were used.

Results

Descriptive Statistics

Our sample consisted of 7056 twins, including 3039 females and 4017 males. The mean age at interview was 36.7 (*SD*=8.82), while the mean age during the heaviest year was 26.1 (*SD*=9). The mean years of education was 13.8 (*SD*=2.5). The distribution of phenotypes in our sample is shown separately for males and females in Table 2.1. There were significant differences in education level, ADsx, and maximum quantity between the sexes. Table 2.2 shows that, in males, there were moderate, significant inverse phenotypic correlations between the educational attainment and the ARPs. In contrast, females showed weaker, non-significant correlations between education and the ARPs. The cross-twin cross-trait correlations for each phenotype were again significant in males, but not in females (Table 2.3). While numerically lower, the MZ correlations were stronger than the DZ correlations in males, suggesting that overlapping genetic influences are likely present between educational qualifications and each alcohol measure. Due to the non-significant cross-

twin cross-trait correlations in females, it is not clear whether genetic influences are shared between education and the ARPs.

Table 2.3. Pearson cross-twin cross-trait correlations between educational attainment and ARPs

	Education-ADsx (95% CI)	Education-Max. Quant. (95% CI)
Monozygotic females	-0.01(-0.11,0.08)	-0.05(-0.14,0.05)
Dizygotic females	0.05 (-0.07,0.18)	0.10 (-0.01,0.22)
Monozygotic males	**-0.15 (-0.22,-0.07)**	**-0.22 (-0.29,-0.15)**
Dizygotic males	**-0.13 (-0.22,-0.05)**	**-0.12 (-0.21,-0.04)**

Notes: ADsx, alcohol dependence symptoms. Significant correlations in bold. Correlations were calculated using the double entry method in which each twin is entered as both Twin 1 and Twin 2.

Table 2.4. Estimates for univariate twin models

	Females			Males		
Trait	A	C	E			
ADsx	.40 (.29-.46)	--	.60 (.53-.68)	.45 (.43-.51)	--	.55 (.49-.61)
Maximum Quantity	.58 (.52-.64)	--	.42 (.36-.47)	.47 (.41-.52)	--	.53 (.48-.59)
Education	.44 (.29-.61)	.31 (.41-.45)	.25 (.22-.39)	.29 (.19-.40)	.49 (.39-.58)	.22 (.19-.25)

Notes: ADsx, alcohol dependence symptoms; A, additive genes; C, common environment; E, unique environment.

Twin Modeling

Comparing the fits of our preliminary univariate saturated and reduced models indicated that the means and variances for each trait could be equated by twin order and zygosity across all phenotypes, but not by sex (results not shown). Thus, we estimated the same means and variances for twin1 and twin2 as well as DZ and MZ twins in subsequent models. Models were run separately for males and females. Univariate estimates for each trait are shown separately by sex in Table 2.4. Our univariate parameter estimates are somewhat different from those of Latvala et al., which is not unusual to find across twin studies, and may be a result of cultural and/or age differences. Nevertheless, they are within the range found by previous studies

Table 2.5. Model fitting results for moderation models in females

Model Description	-2LL	$\Delta\chi^2$	Δdf	*p*-value	ΔAIC
Education/ADsx					
1a.Full	28861.89	--	--	--	--
2a. Drop mod. of common A path	28864.07	2.18	1	0.14	0.18
3a. Drop mod. of unique A path	28864.20	2.31	1	0.13	0.31
4a. Drop mod. of common C path	28863.03	1.13	1	0.29	-0.87
5a. Drop mod. of unique C path	28861.89	0.00	1	1	-2.00
6a. Drop mod. of common E path	28866.30	0.79	1	0.38	-1.21
7a. Drop mod. of unique E path	28861.93	4.41	1	0.04	2.41
Education/Max Quant					
1a.Full	21823.99	--	--	--	--
2a. Drop mod. of common A path	21824.05	0.07	1	0.80	-1.94
3a. Drop mod. of unique A path	21824.62	0.63	1	0.43	-1.37
4a. Drop mod. of common C path	21823.99	0.001	1	0.98	-2.00
5a. Drop mod. of unique C path	21824.19	0.20	1	0.65	-1.80
6a. Drop mod. of common E path	21824.16	0.17	1	0.68	-1.83
7a. Drop mod. of unique E path	21830.26	6.28	1	0.01	4.28

Notes: LL, log likelihood; $\Delta\chi^2$, change in chi-square from full model; *p*-value associated with change in chi-square; ΔAIC, Akaike's criterion from full model. All models compared against full model.

(8,79-81,104).Our moderation model fitting results are shown in Table 2.5 for females and Table 2.6 for males. In females, across both alcohol related phenotypes, moderation of the common and unique paths of the A and C components could be dropped without deterioration in model fit. Also, moderation of the common E component could be dropped, but setting the moderation path for the unique E component was not statistically possible. In males, the same parameters were retained, except in the case of the unique moderation path on A for ADsx. This parameter was border line by *p*-value, but the change in AIC was positive, so we elected to retain it. The graphs of moderation effects shown in Figure 2.3

Table 2.6. Model fitting results for moderation models in males

Model Description	-2LL	$\Delta\chi^2$	Δdf	*p*-value	ΔAIC
Education/ADsx					
1a.Full	28861.89	--	--	--	--
2a. Drop mod. of common A path	28862.285	0.39	1	0.53	-1.61
3a. Drop mod. of unique A path	28866.109	3.68	1	0.06	1.68
4a. Drop mod. of common C path	28866.183	0.70	1	0.40	-1.30
5a. Drop mod. of unique C path	28866.618	1.76	1	0.19	-0.24
6a. Drop mod. of common E path	28905.888	0.23	1	0.63	-1.77
7a. Drop mod. of unique E path	28906.400	23.38	1	0.00	21.38
Education/Max Quant					
1a.Full	21823.99	--	--	--	--
2a. Drop mod. of common A path	21825.03	1.05	1	0.31	-0.95
3a. Drop mod. of unique A path	21824.45	0.47	1	0.50	-1.54
4a. Drop mod. of common C path	21825.04	1.05	1	0.31	-0.95
5a. Drop mod. of unique C path	21824.67	0.69	1	0.41	-1.32
6a. Drop mod. of common E path	21823.99	1.20	1	0.27	-0.81
7a. Drop mod. of unique E path	21852.36	28.38	1	0.00	26.38

Notes: LL, log likelihood; $\Delta\chi^2$, change in chi-square from full model; *p*-value associated with change in chi-square; ΔAIC, Akaike's information criterion from full model. All models compared against full model.

reveal that for both males and females, across the two ARPs, raw E variance decreased with higher levels of education; while raw A variance decreased slightly or remained stable. To compensate for the change in unique environmental variance, the heritability of alcohol related traits was higher in more educated twins.

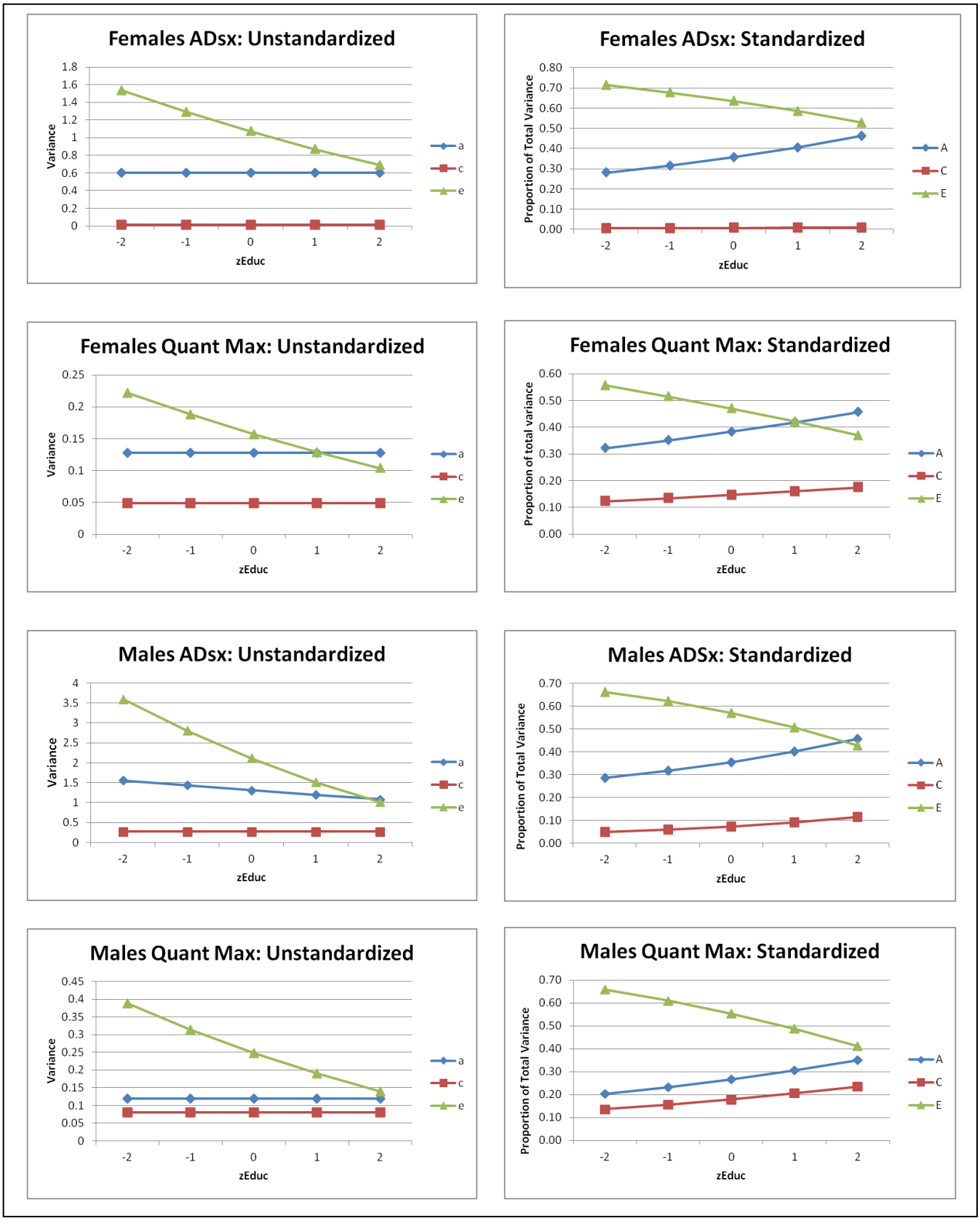

Figure 2.3. Influence of education on variance components in females and males. Unstandardized (a,c,e) and standardized (A,C,E) variance components. First four panels show ADsx and maximum quantity for females; second four show ADsx and maximum quantity for males.

Discussion

Consistent with Latvala and colleagues as well as other studies (74,75,77,104), we found significant negative phenotypic correlations in males between education and ARPs. For females, the correlations were negative but neither was significant, so it is difficult to draw conclusions. However, it is interesting to note that Heurta et al. estimated a positive, significant phenotypic correlation between educational qualifications and frequency of alcohol consumption after controlling for covariates in females but not in males (76). These authors suggest that the positive correlation in females may be a result of greater engagement in traditionally male spheres and greater postponement of childbearing among more highly educated females. Perhaps a more complicated relationship is occurring in our sample of females in which there are direct and inverse correlations between education and the alcohol measures in different individuals that are non-significant when the group is viewed as a whole. Unlike the findings of Latvala and colleagues, the genetic correlations in females were not significant for either ARP.

In contrast, the genetic correlation between education and maximum quantity was significant in males, reflecting the findings of Latvala and colleagues and providing support for overlapping genetic influences between ARPs and educational attainment in this gender. Johnson and colleagues also found a significant negative genetic correlation between education level and an alcohol use composite score consisting of alcohol abuse and dependence symptoms and maximum number of drinks (105). These overlapping genes may act to influence cognitive ability, which affects one's ability to become more highly educated and to make decisions about substance use. Indeed, longitudinal studies have provided evidence that individuals with neurocognitive deficits in childhood and young adulthood show an increased risk for developing alcohol use problems later in life (106). Additionally, in a separate study, Latvala and co-investigators estimated negative genetic correlations between verbal achievement scores and

RAPI, maximum drinks in 24 hours, and AD symptoms (104). Interestingly, other twin studies have demonstrated negative genetic correlations with other substance measures as well as psychiatric phenotypes more generally. McCaffery and co-authors found that the negative phenotypic correlation between smoking initiation and educational qualifications in males was partially explained by overlapping genetic influences (107). Tambs et al. showed that the negative phenotypic correlation between any anxiety disorder and education is mostly explained by overlapping genetic factors in a Norwegian sample (80). Taken together, these studies suggest that there may be a common set of genes that predispose one to lower educational level and psychiatric disorder(s) in general; the form that these psychiatric disorders take may depend on other disorder-specific genes or environmental factors. Additionally, these findings inform the causation debate regarding mental health and socioeconomic status (SES), which is often measured by educational level, occupational status, and income. The debate centers around this question: to what extent do psychiatric disorders lead to lower attainment (social selection) and to what extent does stress associated with lower SES increase propensity to psychiatric disorders (social causation)(80,108)? Generally, investigators agree that both theories together account for the majority of the phenotypic correlation between SES and psychiatric disorders (108). However, the negative genetic correlation in males found in our study and the studies of Latvala and colleagues and Tambs and colleagues suggest that the relationship between SES and anxiety as well as substance use measures is partly explained by overlapping genetic factors, not just by direct causation in either direction.

There was no direct moderation on additive genetic variance for maximum quantity (both in our study and the Finnish twins) and for RAPI in the Finnish twins, whereas there was direct moderation on unique environmental variance for all of these traits such the unique environment

was more influential on variation in ARPs in twins with lower education. Additionally, in the present study, the moderation of additive genetic variance was significant for ADsx, but the change in unique environmental variance (decrease with higher education) was far steeper. To compensate for the change in unique environmental variance, the heritability of the variables in both studies increased such that genetic influences on these traits were relatively more important in those with higher education. Therefore, in most of our results, and all of the results from the Finnish study, only the E x E interaction was significant, and this lead to a heritability x E interaction. It is sensible that there is increased importance of unique environmental influences on variation in alcohol traits in individuals with lower education in that low educational status is associated with stressors that are not as likely to affect those with higher education, such as job instability and limited income (109). It is interesting to note that in the case of ADsx in males, total genetic variance actually decreased slightly; however, to compensate for the much steeper decrease in unique environmental variance, the relative importance of genetic influences increased. The ADsx result in males is a good example of why, although many G x E studies focus on standardized estimates, several authors have pointed out that this may be misleading because raw variance could be changing in the opposite direction (110-113). Higher genetic variance in alcohol related traits among those with less education is reasonable in light of previous work showing that genetic differences among individuals become more influential for externalizing psychopathology among those experiencing greater environmental adversity (114).

Our study has several strengths, including that we tested moderation of genetic and environmental influences that are both common to education and alcohol related traits as well as those unique to ARPs and we considered both standardized and unstandardized variance components. However, we must also consider three methodological limitations. First, we were

not able to use all of the same measures as in the Finnish study because RAPI scores were not assessed in the VATSPSUD; however, these scores were correlated at 0.55 with alcohol dependence symptoms in a subsample of the Finnish twins (73). Second, lifetime symptoms of ADsx and maximum consumption during the heaviest year were based on participants' retrospective recall. It is possible that our results may have been influenced by recall errors. Finally, it is unclear whether our sample of Virginia-born Caucasians generalizes to other populations, although the standardized G x E results appear similar to a Finnish population.

CHAPTER 3: Alcoholic Beverage Preference: An Epidemiological & Twin Study

Abstract

Background: Alcoholic beverage preference (ABP), defined as one's most heavily/frequently consumed alcoholic beverage, has been associated with demographic and lifestyle characteristics, medical outcomes, and drinking habits and their consequences. The present study sought to replicate and extend some of these findings in a twin sample by assessing the risk of endorsing alcohol use disorders and related measures using ABP in both the same individual and the co-twin as the predictor. Additionally, this is the first study to assess the genetic and environmental influences on preference alone as well as those that overlap with total quantity/frequency. Finally, we examined the role of non-additive genetic factors and parental influences on ABP.

Methods: We conducted association analyses in SAS and twin modeling in OpenMx. We used three twin samples, including the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD), FinnTwin12 (FT12), and The Virginia 30,000 (VA30K). Symptoms and diagnoses were based on DSM-IV criteria.

Results: The risk of endorsing AUDs or symptoms, high consumption measures, and other psychiatric conditions was lowest in wine drinkers, except in the case of major depression. Beer drinkers were the most likely to engage in high and frequent consumption, while distilled spirits (DS) drinkers were at the greatest risk for alcohol dependence, AUD symptoms related to impairment, and other psychiatric conditions, except depression. The likelihood of endorsing alcohol abuse and the remaining symptoms of AUDs was not different between beer and DS preference drinkers. Furthermore, across all three twin samples, the genetic overlap between total

consumption measures and preference was significant, but only small to moderate, and two out of the three twin samples showed that additive genetic influences on preference were greater in females than males. Finally, non-additive genetic factors and parental influences do not appear to be important sources of variation for ABP.

Conclusions: Preference for particular alcoholic beverages is associated with differential risk for high/frequent consumption and the development of AUDs and other psychiatric conditions, suggesting that preference may be a helpful variable in defining subtypes of alcohol dependent individuals. Furthermore, that beer carries the same risk as spirits for AA and most symptoms of AUDs has implications for public health efforts. Finally, our genetic findings may be informative for gene finding efforts of AUDs.

Introduction

Although definitions in the extant epidemiological literature vary, alcoholic beverage preference (ABP) is generally defined as one's most heavily or frequently consumed alcoholic beverage. Previous studies have examined the relationship between preferred alcoholic beverage and demographic and lifestyle characteristics, morbidity and mortality, and drinking habits and their consequences. Wine preference drinkers tend to eat a healthier diet, exercise more frequently, and have a lower incidence of several medical conditions than drinkers who prefer beer or distilled spirits (DS) (115-119). Wine drinkers also consume the least alcohol with the lowest frequency and are the least susceptible to alcohol-related problems and disorders (120-126). In studies that find correlations between ABP and particular health outcomes, such as stroke, obesity, and cancer, the authors suggest the possibility that particular alcohol concentrations and/or beverage-specific bioactive non-ethanol substances (congeners) may contribute to effects on health (127-129). The average ethanol concentration in alcoholic

beverages varies widely across the main classes: beer ≈ 4.5%, wine ≈ 12.9%, and distilled spirits ≈ 41.1% (130). Some studies that find beverages with higher concentrations of ethanol increase the risk for digestive tract cancers suggest that exposure of digestive mucosa to concentrated ethanol may be at play (131,132). Coder et al. proposed that spirits drinkers are the most prone to alcohol-related disease because they are able to drink higher amounts more quickly due to the decreased volume of liquid compared to other alcoholic beverages (133). Furthermore, wine, beer, and spirits contain varying levels of congeners and contaminants, including polyphenols, ethyl carbamate, and acetylaldehyde, shown to have direct effects *in vitro* and *in vivo* (134-136). Polyphenols, especially resveratrol in red wine, have been studied extensively for their potential beneficial cardiovascular and anti-carcinogenic effects. Several studies suggest that they have anti-coagulatory, anti-inflammatory, and anti-proliferative effects independent of ethanol (137).

Despite the differential associations between particular forms of ethanol and lifestyle, demographic, and health characteristics, few epidemiological studies consider them separately or control for differences in analyses with aggregate consumption measures. Furthermore, to our knowledge, no studies have specifically examined the heritability of preference for particular alcoholic beverages through comparison of intake of one type of beverage to the others. However, two groups have assessed the genetic and environmental influences on intake frequency of specific alcoholic beverages. Kaprio and colleagues (138,139) found the heritabilities for frequency of beer (and spirits) intake in females and males, respectively, to be 0.46 and 0.39 (0.36 and 0.38). Fabsitz and colleagues (140) presented twin correlations only, but if heritabilities were estimated from these correlations using Falconer's broad sense formula (141), they would be as follows: 0.22 for beer, 0.15 for wine, and 0.44 for other (not beer or wine).

The first aim of our study was to replicate the associations that other investigators have found between ABP (assessed from data about the frequency of consumption of wine, beer, and spirits) and alcohol consumption (AC) measures in one of our twin samples. Second, we sought to examine relationships that have received less attention in the literature, including between ABP and AUD symptoms and diagnoses, major depression (MD), antisocial personality disorder (ASPD), and any symptom of illicit substance use disorders. Third, to provide a sense of the degree to which familial (i.e. genetic and common environmental) factors contribute to the associations between measures, we predicted the odds of reporting all measures in one twin based on the co-twin's beverage preference. Our fourth goal was to evaluate the genetic and environmental underpinnings of the association between frequency of total consumption and ABP, as well as the unique influences on preference, in the same sample that was used in the previous three aims. Finally, we used two additional twin samples with continuous consumption measures of specific alcoholic beverages to answer the following questions: (1) Will the findings from the first twin sample reflect those in the second two? (2) What role do non-additive genetic factors and parental influences play in ABP?

Methods

VATSPSUD

To accomplish the first four aims, we used data from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD), a longitudinal study of psychiatric and substance-related conditions in 2 samples of Caucasian adult twins identified through the population-based Virginia Twin Registry. One sample is comprised of female-female twin pairs, while the other includes both male-male and male-female twins. For further details of recruitment and subject characteristics, see Kendler and Prescott (91). We utilized data for the

present report from the fourth wave of the female-female study (FF4) and the second wave of male-male and male-female twins (MM/MF2) because these were the only waves that included the ABP question. After receiving a description of the protocol, participants provided written consent for in-person interviews and verbal consent for telephone interviews. The Office of Research Subjects Protection at Virginia Commonwealth University approved this project. Zygosity was determined by a computer algorithm of standard questionnaire responses validated against DNA genotyping of 496 twin pairs. A total of 5,489 individual twins with non-missing data for the preference question were used in the regression analyses. Fewer twins (N=3,541) were part of the twin models because we excluded opposite sex pairs for simplicity.

Clinically trained interviewers assessed ABP by asking "During the year when you drank the most, what type of alcoholic beverage did you drink most often?" The DS category included vodka, gin, rum, bourbon, scotch, and liqueurs. Due to the low proportion of individuals who endorsed malt liquor as their beverage of preference (N=15, 0.27%), we did not include this group of subjects in our analyses. If a subject indicated two beverages were consumed with equal frequency, he was categorized according to the first beverage mentioned. The interviewers also asked about AC measures in the heaviest year, including drink frequency, regular quantity, maximum quantity in 24 hours, and frequency of intoxication (drunk frequency). For the latter measure, we had 11 response options: 1 = Never, 2 = Once/Year, 3 = 2 Times/Year, 4 = 3 to 6 Times/Year, 5=7 to 11 Times⁄ Year, 6 = 1 Time/Month, 7 = 2 to 3 Times/ Month, 8 = 1 to 2 Times/Week, 9 = 3 to 4 Times/Week, 10 = 5 to 6 Times/Week, and 11 = Every Day. Due to the skewed distributions of the first three continuous AC measures and the unequal distribution of individuals in categories for drunk frequency, we divided each measure into five roughly equal ordinal categories. There were different cutoffs for females and males due to the differences in

their distributions. Respectively, the cutoffs for females and males were as follows: drink frequency (1, 2-3, 4-7, 8-14, 15-30; 1-3, 4-9, 10-15, 16-27, 28-30 days/mo.), regular quantity (1, 2, 3, 4, ≥5; 1-2, 3, 4-5, 6-8, ≥9 drinks/day), maximum quantity (1, 2-3, 4, 5-6, ≥7; 1-5, 6-9, 10-12, 13-20, ≥21 drinks/day), and, using the response options listed above, drunk frequency (1, 2, 3 to 4, 5-7, and 8-11; 1- 2, 3- 5, 6- 7, 8, and 9-11). The twins were informed that one drink of beer (12oz), wine (5oz), or spirits (1.5oz) equated to a standard alcoholic "drink". Additionally, the interviewers used an adaptation of the Structured Clinical Interview for DSM-IV (6,92) to inquire about lifetime symptoms of AUD, MD, ASPD, and illicit substance abuse and dependence; lifetime diagnoses were determined algorithmically using symptom data.

FinnTwin12

FinnTwin12 (FT12) is a population-based longitudinal study of twins born in Finland between 1983 and 1987. Because twins were ascertained through the Central Population Registry, which includes records of every citizen, FT12 is an unbiased sample. Twin participants (N=1,854 individual twins) were interviewed at different ages using the Semi-Structured Assessment for the Genetics of Alcoholism (142), a psychiatric interview developed for COGA. Additional details of study design can be found elsewhere (143). To assess quantity of intake of particular beverages, subjects were asked about their consumption of wine, beer, spirits in a typical week. Although there is also phenotypic data available for ages 12, 14, and 17, we chose to focus on age 22 because a stable preference is more likely to be present at this age and this age group would have the most access to all types of alcoholic beverages. In Finland, one must be 18 to buy alcoholic beverages containing ≤ 22% alcohol by volume and 20 to purchase beverages greater than that (144). Although prior to the 1950's Finland's alcohol intake was dominated by

the consumption of spirits, beer consumption has increased dramatically to become the most consumed alcoholic beverage and wine consumption has increased steadily, now accounting for 16% of total intake (145). We believe that the availability of all three beverages is adequate to assume that amount consumed is reflective of preference. As in studies of coffee preference (146,147), we defined ABP as the ratio of a particular alcoholic beverage to the total quantity consumed. Due to the skewed nature of the distributions for total quantity and preference of wine, beer, and spirits, we polychotomized these variables into three roughly equal categories with separate cutoffs for females and males.

VA30K

The Virginia 30,000 (VA30K) sample contains data from 14,763 individual twins, along with additional relatives, ascertained from two sources, including Virginia public records and through response to a newsletter published by the American Association of Retired Persons (148). Twins participating in the study were mailed a Health and Lifestyles questionnaire (HLS) and were asked to supply the names and addresses of their spouses, siblings, parents, and children for the follow-up study of relatives of twins. Zygosity of twins was assigned based on responses to standard questions about similarity and the degree to which others confused them. For the purpose of this report, we used a question from the Drinking Habits section of the HLS: "Please describe your consumption of alcohol LAST WEEK." Participants were provided with a chart on which they could record their consumption of beer, wine, and liquor on each day of last week. Although alcohol consumption in a typical week would be more informative and similar to the phenotype measured in FT12, unfortunately, this data is not available in the VA30K. As with FT12, we defined ABP as the ratio of a particular alcoholic beverage to the total quantity

consumed and polychotomized preference and total consumption into three roughly equal categories with separate cutoffs for females and males. It should be noted that there is some overlap between VA30K and the VATSPSUD samples.

Multivariate Regression in VATSPSUD

To accomplish aims 1-4, we estimated the odds of endorsing the highest level of alcohol consumption compared to the lowest level, AUDs and their individual symptoms, and other psychiatric disorders using ABP as the determinant. If the within-person odds ratios (ORs) were significant, we also tested for associations between these variables in one twin using the beverage preference of the co-twin to assess for common familial influences between ABP and the other phenotypes. We utilized the GENMOD procedure in SAS 9.2 (93) to account for the correlated nature of twin data, and sex, zygosity, age at interview, age during heaviest year, years of education, lifetime smoking status, and current marital status as covariates.

Twin Resemblance & Bivariate Model Fitting in the VATSPSUD

Because ABP does not have an underlying continuous distribution, we analyzed each class of beverage separately as a binary variable, categorizing individuals based on whether they preferred that beverage or not. We then calculated tetrachoric correlations and 95% confidence intervals for ABP in MZ and DZ twin pairs using SAS 9.2 (93). Because our co-twin regression analyses of AC measures in the VATSPSUD indicated some overlap in familial factors contributing to drink frequency and preference, we used bivariate Cholesky models (Figure 3.1) to evaluate (1) the extent to which genetic and environmental influences contribute to the covariation between total frequency/quantity and preference and (2) the contribution of these influences independently on preference across all three twin samples. We performed all model fitting within

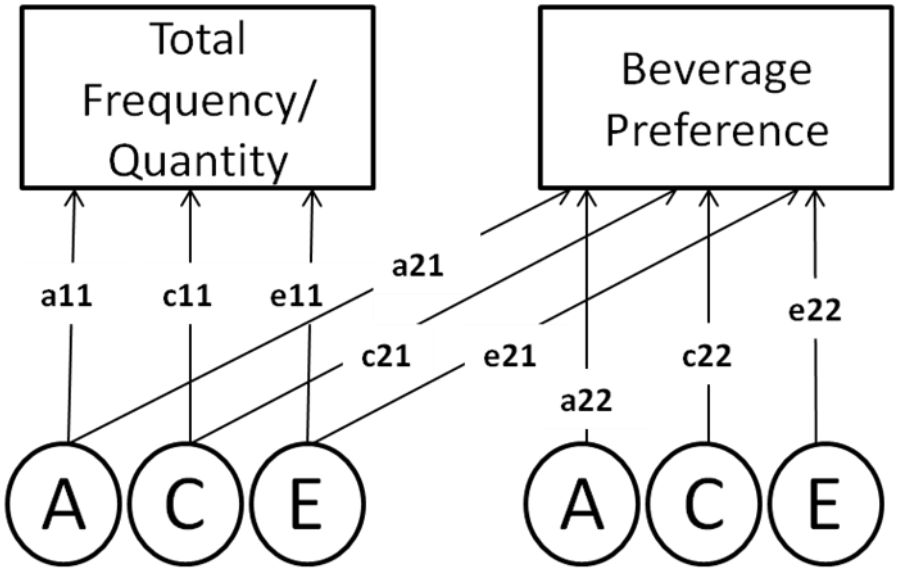

Figure 3.1. Bivariate Cholesky decomposition. A, additive genetic; C=shared environment; E, unique environment; a11, genetic path for total frequency/quantity; c11, shared environmental path for total frequency/quantity; e11, unique environmental path for total frequency/quantity; a12, genetic covariance between total frequency/quantity and preference; c12, shared environmental covariance between total frequency/quantity and preference; e12, unique environmental covariance between total frequency/quantity and preference; a22, genetic path unique to preference; c22, shared environmental path unique to preference; e22, unique environmental path unique to preference.

the SEM program OpenMx 1.1 (149) using four groups, including MZ and DZ twins of both sexes. We did not use opposite sex twins due to the added complexity of including them in the twin models. Details describing the classic twin design and the bivariate model fitting procedure can be found in the previous chapter.

Extended Twin Family Design (ETFD)

By using relatives in addition to MZ and DZ twins, researchers are able to estimate sources of phenotypic variation beyond A, C, and E. The most extensive model involves utilizing information from twins, parents, siblings, spouses, and offspring to parse variation into two sex-specific genetic components, including additive (A) and non-additive (D),

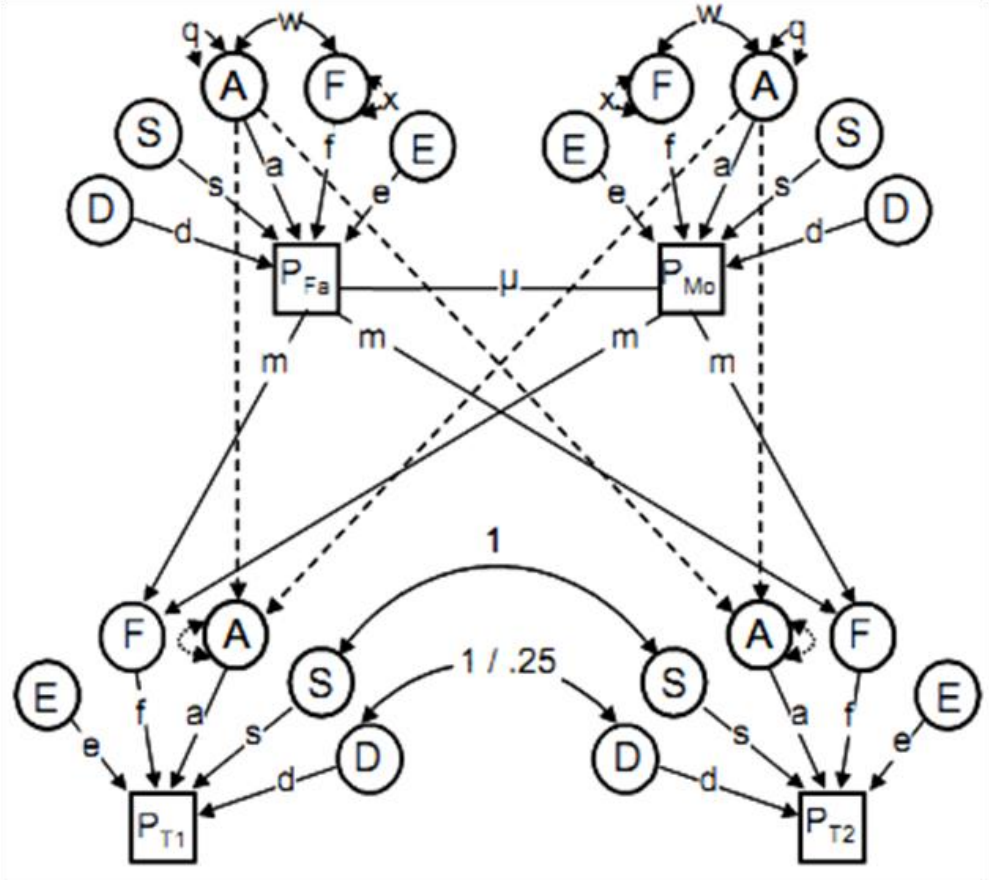

Figure 3.2. Cascade model for extended twin designs. Adapted from Keller et al., 2009 (43)

and four sex-specific environmental components, including common sibling (S), special twin (T), vertical cultural transmission (F), and unique environment (E). D refers to interactions between alleles within a locus (dominance) or across different loci (epistasis); dominance and epistasis cannot be distinguished in this model (150). S, T, and F are all components of the shared environment. S is the non-parental component shared between all offspring, while T represents an additional environmental correlation that makes twins more similar than siblings (97); it is attributed to sharing the same intrauterine environment and being the same age. F refers to the proportion of the shared environment that is transmitted from parents to offspring (151). The most recent implementation of the ETFD is the *Cascade* model (Figure 3.2) (152), which we used to conduct univariate analyses of preference for wine, beer, and spirits. Again, we performed all model fitting with OpenMx 1.1 (149). This time we included opposite sex twins, allowing us to test for both qualitative (i.e. different genes and/or environments) and quantitative (i.e. different magnitude of effects) sex differences. We fitted the full ASETFD *Cascade* model, which may include parameter estimates that are not significantly different from zero, to avoid the parameter biasing that can occur when parameters are dropped from the model (153).

Results

Descriptive Statistics in the VATSPSUD

Descriptive statistics by preference are summarized in Table 3.1. Wine preference drinkers were more likely to be female (83.1%), more educated (mean=14.8 years, SD=2.3), older during their heaviest year of drinking (mean=34.7 years, SD=10.3), and married (64.7%) compared to the other two classes. DS drinkers were split fairly evenly across the sexes (female=53.6%) and fell between the other groups in terms of age (mean=28.6 years, SD=9.8). They were also the least educated (mean=13.7 years, SD=2.5) of all drinking classes. Finally,

beer drinkers were most likely to be male (69.4%) and unmarried (42.8%) and were, on average, the youngest (mean=25.4 years, SD=8.4) group of preference drinkers. Our demographic findings were generally consistent with previous work (154-156).

Table 3.1. Descriptive statistics in the VATSPSUD

	Drinking Preference in Heaviest Year		
Variable	Wine N= 584	Beer N=3578	Distilled Spirits N=1327
Sex [a], N (%) Female Male	 485 (83.1) 99 (16.9)	 1093 (30.6) 2485 (69.4)	 711 (53.6) 616 (46.4)
Smoking Status [a,b], N (%) Regular Smoker Never Smoked	 186 (50.4) 183 (49.6)	 2144 (69.8) 926 (30.2)	 785 (74.4) 270 (25.6)
Zygosity [a,b], N (%) MZ DZ	 241 (41.3) 343 (58.7)	 1348 (37.7) 2230 (62.3)	 468 (35.3) 859 (64.7)
Heaviest year age [a], mean (SD)	34.7 (10.3)	25.4 (8.4)	28.6 (9.8)
Education [a], mean (SD)	14.8 (2.3)	13.9 (2.5)	13.7 (2.5)
Marriage Status [a], N (%) Married Unmarried	 378 (64.7) 206 (35.3)	 2046 (57.2) 1532 (42.8)	 807 (60.8) 520 (39.2)

Notes: [a] Significant difference at $p < 0.05$ using χ^2 for categorical variables and ANOVA for continuous variables. [b] Some data is missing for this variable.

Multivariate Regression in the VATSPSUD

Using wine as the reference category, Table 3.2 shows the ORs for the highest level of consumption across the four AC measures compared to the first level. Those who preferred spirits were up to ~5 times more likely to endorse the highest levels of consumption compared to wine drinkers, while those who prefer beer were up to ~9 times more likely. Additionally, when beer drinkers were compared to DS drinkers, the risk the highest level of drink and drunk frequency were significantly higher for beer drinkers. Most co-twin estimates were significant,

suggesting that common familial factors contribute to the phenotypic association between AC measures and ABP.

Table 3.2. Association of ABP with AC measures in the VATSPSUD

		ABP in Heaviest Year (Wine Reference)		ABP in Heaviest Year (Beer Reference)
Consumption Variable		Beer Adjusted OR[a] (95% CI)	Distilled Spirits Adjusted OR[a] (95% CI)	Distilled Spirits Adjusted OR[a] (95% CI)
Drink Frequency				
Female	Male			
1 day/mo.	1-3 day/mo.	--	--	--
15-30 day/mo.	28-30 day/mo.	**3.18 (2.07-4.88)**	1.32(0.86-2.03)	**0.42(0.32-0.55)**
Co-twin	--	1.41 (0.92-2.15)	--	0.81(0.62-1.06)
Regular Quantity				
Female	Male			
1 drink/wk	1-2 drink/wk	--	--	--
≥5 drink/wk	≥9 drink/wk	**6.40 (3.74-10.97)**	**4.95 (2.85-8.63)**	0.78 (0.60-1.03)
	--	**2.27 (1.42-3.62)**	**1.96 (1.19-3.21)**	--
Maximum Quantity				
Female	Male			
1 drink/wk	1-5 drink/wk	--	--	--
>7 drink/wk	>21 drink/wk	**8.99 (5.22-15.46)**	**4.03 (2.29-7.07)**	**0.46 (0.34-0.62)**
Co-twin	--	**2.13 (1.32-3.45)**	1.48 (0.89-2.47)	**0.69 (0.50-0.96)**
Drunk Frequency				
Female	Male			
1	1-2	--	--	--
8-11	9-11	**5.30 (2.99-9.39)**	**4.05 (2.23-7.36**	0.78 (0.57-1.07)
Co-twin	--	**1.94 (1.23-3.06)**	1.53 (0.95-2.46)	--

Notes: Wine preference drinkers are reference group in first set of analyses and beer preference drinkers are reference group in second set of analyses. ABP, alcoholic beverage preference; OR, Odds Ratio; CI, Confidence Interval. [a]Multivariate logistic regression accounting for continuous (age at interview, age during heaviest year, years of education) and binary (sex, zygosity, lifetime smoking status, and current marital status) covariates.

Table 3.3 shows that the odds of reporting an AUD or the majority of symptoms of these disorders are significantly higher in those who prefer beer or DS as compared to those who prefer wine. When DS drinkers were compared to beer preference drinkers, estimates were significantly higher for the former vs. the latter for endorsing AD and certain symptoms of

AUDs, including withdrawal, preoccupation with obtaining, using or recovering from the effects of alcohol, interference with daily activities, and use despite physical/psychological problems.

Table 3.3. Association of ABP with lifetime DSM-IV AUD sxs/diagnoses in the VATSPSUD

	ABP in Heaviest Year (Wine Reference)		ABP in Heaviest Year (Beer Reference)
Lifetime DSM- IV AAsx/ADsx	Beer OR (95% CI)	Distilled Spirits OR (95% CI)	Distilled Spirits OR (95% CI)
AA Co-twin	**3.63 (2.36-5.59)** 1.07 (0.80-1.45)	**3.51 (2.26-5.45)** 1.06 (0.77-1.45)	0.97 (0.83-1.14) --
Failure to fulfill major obligations Co-twin	**4.76 (2.39-9.48)** 1.06 (0.72-1.55)	**5.20 (2.59-10.43)** 1.18 (0.79-1.75)	1.10 (0.90-1.33)
Physically hazardous use Co-twin	**3.21 (1.79-5.75)** 1.17 (0.82-1.67)	**3.44 (1.90-6.23)** 1.17 (0.81-1.71)	1.08 (0.90-1.30) --
Use leading to legal problems Co-twin	**3.18 (1.54-6.57)** **2.10 (1.20-3.69)**	**2.76 (1.31-5.80)** **1.98 (1.1-3.53)**	0.88 (0.70-1.10) --
Use despite social problems Co-twin	**2.61 (1.64-4.16)** **1.57 (1.07-2.32)**	**2.79 (1.74-4.47)** **1.67 (1.12-2.50)**	1.07 (0.89-1.28) --
AD Co-twin	**2.31 (1.43-3.73)** 1.28 (0.89-1.85)	**2.96 (1.82-4.81)** 1.30 (0.88-1.92)	**1.28 (1.08-1.53)** 1.03 (0.84-1.27)
Tolerance Co-twin	**2.62 (1.60-4.29)** 1.09(0.77-1.55)	**2.99 (1.80-4.97)** 1.04 (0.71-1.51)	1.14 (0.96-1.37) --
Withdrawal Co-twin	**2.05 (1.03-4.07)** 1.44 (0.83-2.50)	**2.76 (1.38-5.53)** 1.59 (0.89-2.82)	**1.35 (1.07-1.71)** 1.12 (0.85-1.48)
Consumed more than intended Co-twin	**2.59 (1.77-3.79)** 1.17 (0.86-1.59)	**2.56 (1.73-3.78)** 1.09 (0.79-1.50)	0.99 (0.84-1.16) --
Unsuccessful attempts to cut down Co-twin	**1.97 (1.30-2.98)** 1.03 (0.74-1.46)	**2.09 (1.37-3.19)** 1.08 (0.76-1.53)	1.06 (0.89-1.27) --
Preoccupation Co-twin	**2.05 (1.29-3.26)** 1.21 (0.84-1.76)	**2.55 (1.60-4.07)** 1.24 (0.85-1.83)	**1.24 (1.03-1.48)** 1.04 (0.84-1.28)
Interference Co-twin	**2.54 (1.03-6.27)** 1.18 (0.67-2.09)	**4.69 (1.90-11.61)** 1.14 (0.62-2.08)	**1.85 (1.44-2.39)** 0.98 (0.72-1.34)
Use despite physical/psych.problems Co-twin	1.41 (0.48-4.12) --	2.78 (0.94-8.21) --	**2.00 (1.41-2.83)** 1.28 (0.86-1.91)

Notes: Wine preference drinkers are reference group in first set of analyses and beer preference drinkers are reference group in second set of analyses. ABP, alcoholic beverage preference; OR= Odds Ratio; CI, Confidence Interval. [a]Multivariate logistic regression accounting for continuous (age at interview, age during heaviest year, years of education) and binary (sex, zygosity, lifetime smoking status, and current marital status) covariates. Estimates significant at $p< 0.05$ are bolded.

However, there was no difference in risk between beer and DS preference drinkers for alcohol abuse and the remaining symptoms of both disorders. When the likelihood for endorsement of AA, AD, and AUD symptoms was predicted from the beverage preference of the co-twin, only two estimates were significant.

TABLE 3.4. Association of ABP with lifetime DSM-IV sxs/diagnoses in the VATSPSUD

	ABP in Heaviest Year (Wine Reference)		ABP in Heaviest Year (Beer Reference)
DSM- IV Diagnosis	Beer Adjusted OR[a] (95% CI)	Distilled Spirits Adjusted OR[a] (95% CI)	Distilled Spirits Adjusted OR[a] (95% CI)
MD Co-twin	**0.71 (0.55-0.92)** 0.82 (0.61-1.09)	1.14 (0.88-1.49) --	**1.60 (1.37-1.87)** **1.23 (1.03-1.47)**
ASPD Co-twin	1.74 (0.95-3.20) --	**2.11 (1.12-3.96)** 1.33 (0.81-2.19)	1.21 (0.96-1.52) --
Any DAsx Co-twin	**1.50 (1.01-2.21)** 1.18 (0.82-1.70)	**2.13 (1.43-3.17)** 1.27 (0.87-1.85)	**1.42 (1.20-1.69)** 1.09 (0.89-1.34)
Any DDsx Co-twin	1.08 (0.61-1.90) --	**1.77 (1.00-3.14)** 1.22 (0.72-2.07)	**1.63 (1.28-2.08)** 1.13 (0.85-1.52)

Notes: Wine preference drinkers are reference group and beer preference drinkers are reference group in second set of analyses. ABP=alcoholic beverage preference; OR, Odds Ratio; CI, Confidence Interval. [a]Multivariate logistic regression accounting for continuous (age at interview, age during heaviest year, years of education) and binary (sex, zygosity, lifetime smoking status, and current marital status) covariates. Estimates significant at $p < 0.05$ are bolded.

Table 3.4 lists the likelihood of endorsing MD, ASPD, and any symptom of illicit substance abuse/dependence across each drinking class. Compared to those who prefer wine, beer drinkers were at higher risk for any symptom of substance abuse, but for no other symptom/diagnosis, and DS preference drinkers were at the greater risk for ASPD, and any symptom of illicit substance abuse or dependence, but not MD. Interestingly, wine preference drinkers were at a significantly higher risk for depression as compared to beer drinkers. When beer was used as the reference category, MD and the substance abuse and dependence estimates were significantly greater for DS, but ASPD was not. Only one co-twin estimate was significant. Due to the phenotypic associations between ABP and consumption found in this report and

others, we also ran the multivariate logistic regression analyses shown in Table 3.3 and 3.4 using regular quantity as a covariate in addition to the other previously utilized covariates (data not shown). While the estimates declined, most remained significant.

Table 3.5. Twin correlations for ABP across 3 alcoholic beverage types in the VATSPSUD

Sex and zygosity	Number of 'complete' pairs	Total # of Individuals	Tetrachoric Twin Correlations (95% CI)		
			Wine	Beer	DS
Female MZ twins	292	815	.52 (.35-.68)	.49 (.34-.64)	.42 (.24-.58)
Female DZ twins	179	538	.30 (.01-.58)	.28 (.06-.50)	.32 (.10-.54)
Male MZ twins	456	1242	.52 (.20-.84)	.46 (.32-.61)	.36 (.19-.53)
Male DZ twins	293	946	.69 (.39-.99)	.42 (.23-.60)	.33 (.12-.54)
Total	1220	3541			

Notes: ABP, alcoholic beverage preference; CI, Confidence Interval. All correlations are significant.

Twin Correlations & Bivariate Twin Model Fitting in VATSPSUD

Table 3.5 shows that, across all three beverages, the tetrachoric correlations for female MZ twins were greater than female DZ twins, which is consistent with a genetic influence on ABP in females. The difference between male MZ and DZ correlations, by contrast, is much less than for the females, suggesting a smaller genetic influence on preference in males. Note male wine estimates show higher DZ than MZ correlations, reflecting the imprecision of the estimates due to the low number of males preferring wine. There is no reason to expect that DZ twins would be more correlated that MZ twins for any trait.

The phenotypic correlations between drink frequency and the three preference measures varied widely across preferences and between the sexes. For wine, beer, and DS preferences, respectively, the correlations for females (and males) were -.28, .20, -.19 (-.32, .62, .02). Across the three beverages, we could equate thresholds by twin order and zygosity but not by sex. Therefore, we estimated thresholds separately by sex in our models. Bivariate model fitting results in the VATSPSUD are shown in Table 3.6. We found that the models in which the parameter estimates were equated across the sexes (Model 2a-c) fit significantly worse by ΔAIC for all three preferences than the models in which the estimates were allowed to vary by sex (Model 1a-c). Therefore, we estimated all parameters separately for males and females.

Table 3.6. Model fitting results for regular frequency/ABP in the VATSPSUD

Model	Model Comparison	-2LL	$\Delta\chi^2$	Δdf	*p*-value	ΔAIC
Regular Quantity/Wine Preference						
1a. AfCfEf ≠AmCmEm, *ra, rc, re*	-	25157.5	-	-	-	-
2a. AfCfEf =AmCmEm, *ra, rc, re*	2a v. 1a	25177.1	19.6	9	0.02	1.6
3a. CfEf ≠CmEm, *ra, rc, re*	3a v. 1a	25221.3	63.8	6	0.00	51.8
4a. AfEf ≠AmEm, *ra, rc, re*	4a v. 1a	25245.7	88.2	6	0.00	76.2
5a. AfCfEf ≠AmCmEm	5a v. 1a	25283.5	126.0	6	0.00	114.0
Regular Quantity/Beer Preference						
1b. AfCfEf ≠AmCmEm, *ra, rc, re*	-	28376.5	-	-	-	-
2b. AfCfEf =AmCmEm, *ra, rc, re*	2b v. 1b	28465.3	88.8	9	0.00	70.8
3b. CfEf ≠CmEm, *ra, rc, re*	3b v. 1b	28441.0	64.5	6	0.00	52.5
4b. AfEf ≠AmEm, *ra, rc, re*	4b v. 1b	28493.3	116.8	6	0.00	104.8
5b. AfCfEf ≠AmCmEm	5b v. 1b	28564.3	187.8	6	0.00	175.8
Regular Quantity/Spirits Preference						
1c. AfCfEf ≠AmCmEm, *ra, rc, re*	-	27961.2	-	-	-	-
2c. AfCfEf =AmCmEm, *ra, rc, re*	2c v. 1c	28013.9	52.8	9	0.00	34.7
3c. CfEf ≠CmEm, *ra, rc, re*	3c v. 1c	28043.0	81.8	6	0.00	69.8
4c. AfEf ≠AmEm, *ra, rc, re*	4c v. 1c	27978.3	17.1	6	0.01	5.1
5c. AfCfEf ≠AmCmEm	5c v. 1c	28007.8	46.6	6	0.00	34.6

Notes: ABP, alcoholic beverage preference; Af/m, additive genes in females/males; Cf/m, common environment in females/males; Ef/m, unique environment in females/males; LL, log likelihood; $\Delta\chi^2$, change in chi-square from full model; *p*-value associated with change in chi-square; ΔAIC, Akaike's criterion from full model; *ra*, genetic correlation; *rc*, shared environmental correlation; *re*, unique environmental correlation.

When we tested the more parsimonious models AE and CE across both sexes (Models 3a-c & 4a-c), we could drop neither A nor C for any bivariate model. We next found that the genetic (r_a), common environmental (r_c), and unique environmental (r_e) correlations between the phenotypes could not be dropped simultaneously (Model 5a-c).

TABLE 3.7. Parameter estimates for bivariate Cholesky decomposition for each ABP and liability shared with regular quantity in the VATSPSUD

		Females			Males		
		Estimate for Preference	With Regular Frequency	Unique	Estimate for Preference	With Regular Frequency	Unique
Wine	A	47.9	5	95	8.5	100	0
	C	1.0	89	11	49.9	15	85
	E	51.1	12	88	41.6	2	98
Beer	A	25.7	21	79	10.1	8	92
	C	20.2	14	86	36.1	8	92
	E	54.1	13	87	53.7	0	100
Distilled Spirits	A	41.1	30	70	8.6	4	96
	C	2.7	96	4	28.1	8	92
	E	56.3	1	99	63.4	4	96

Notes: ABP, alcoholic beverage preference; A, proportion of variance due to additive genetic effects; C, proportion due to common environment; E, proportion due to unique environment.

In Table 3.7, we show the amount of variance contributed to the three preference phenotypes by A, C, and E. We do not present parameters estimates for drink frequency separately because these were previously assessed in this twin sample (9). The estimates for preference showed a moderately heritable component in females (A_{wine} = 47.9%, A_{beer} =25.7%, A_{DS} = 41.1%) and a smaller heritable component in males (A_{wine} =8.5%, A_{beer} = 10.1%, A_{DS} = 8.6%). These estimates also suggested little to modest common environmental influence in females (C_{wine} =1.0%, C_{beer} = 20.2%, C_{DS} = 2.7%) and a stronger common environmental influence in males (C_{wine} = 49.9%, C_{beer} =36.1%, C_{DS} =28.1%). With the exception of wine preference in males, which has the lowest heritability, the range in overlap of genetic factors

between regular frequency and preference across the three drink categories was small to modest (4% to 30%).

Table 3.8. Model fitting results for total quantity in a typical week/ABP in FT12

Model	Model Comparison	-2LL	$\Delta\chi^2$	Δdf	*p*-value	ΔAIC
Regular Quantity/Wine Preference						
1a. AfCfEf ≠AmCmEm, *ra, rc, re*	-	2844.8	-	-	-	-
2a. AfCfEf =AmCmEm, *ra, rc, re*	2a v. 1a	2862.1	17.3	9	0.04	-0.7
3a. CfEf ≠CmEm, *ra, rc, re*	3a v. 1a	2860.9	16.1	6	0.01	4.1
4a. AfEf ≠AmEm, *ra, rc, re*	4a v. 1a	2852.1	7.3	6	0.29	-4.7
5a. AfCfEf ≠AmCmEm	5a v. 4a	3942.0	1097.2	10	0.00	1081.9
Regular Quantity/Beer Preference						
1b. AfCfEf ≠AmCmEm, *ra, rc, re*	-	3435.8	-	-	-	-
2b. AfCfEf =AmCmEm, *ra, rc, re*	2b v. 1b	3486.6	50.8	9	0.00	32.8
3b. CfEf ≠CmEm, *ra, rc, re*	3b v. 1b	3448.1	12.3	6	0.06	0.3
4b. AfEf ≠AmEm, *ra, rc, re*	4b v. 1b	3438.3	2.5	6	0.87	-9.6
5b. AfCfEf ≠AmCmEm	5b v. 4b	14850.5	11414.7	10	0.00	11404.3
Regular Quantity/Spirits Preference						
1c. AfCfEf ≠AmCmEm, *ra, rc, re*	-	3347.1	-	-	-	-
2c. AfCfEf =AmCmEm, *ra, rc, re*	2c v. 1c	3366.7	19.6	9	0.02	1.6
3c. CfEf ≠CmEm, *ra, rc, re*	3c v. 1c	3363.3	16.2	6	0.01	4.2
4c. AfEf ≠AmEm, *ra, rc, re*	4c v. 1c	3352.1	5.0	6	0.55	-7.1
5c. AfCfEf ≠AmCmEm	5c v. 4c	5876.3	2529.1	10	0.00	2516.2

Notes: ABP, alcoholic beverage preference; Af/m, additive genes in females/males; Cf/m, common environment in females/males; Ef/m, unique environment in females/males; LL, log likelihood; $\Delta\chi^2$, change in chi-square from full model; *p*-value associated with change in chi-square; ΔAIC , Akaike's criterion from full model; *ra*, genetic correlation; *rc*, shared environmental correlation; *re*, unique environmental correlation.

Bivariate Twin Model Fitting in FT12 & VA30K

For the FT12 bivariate models, Table 3.8 shows that the parameter estimates cannot be equated by sex, similar to the results in the VATSPSUD. Therefore, we again estimated all parameters separately for each sex. Dropping A in the three separate analyses (Models 3a-c) resulted in significantly worse fit by ΔAIC and *p*-value except in the case of beer, which was borderline (ΔAIC=0.3; *p*=0.06). We found that we could set C to zero across all three models (Models 4a-c). Thus, in Table 3.9 we present sex-specific AE models. Next, we found that the additive genetic and unique environmental correlations between total quantity and each of the preferences could be not dropped simultaneously without a significant decrease in fit by ΔAIC (Models 5a-5c).

Table 3.9. Parameter estimates for bivariate Cholesky decomposition for ABP and liability shared with total quantity in a typical week in FT12

		Females			Males		
		Estimate for Preference	With Regular Quantity	Unique	Estimate for Preference	With Regular Quantity	Unique
Wine Preference	A	56.4	2	98	21.3	4	96
	E	43.6	19	81	78.7	1	99
Beer Preference	A	70.3	28	72	45.2	0	100
	E	29.7	2	98	54.8	10	90
Spirits Preference	A	44.5	21	79	19.1	31	69
	E	55.5	14	86	80.9	7	93

Notes: ABP, alcoholic beverage preference; A , proportion of variance due to additive genetic effects; E, proportion due to unique environment.

Table 3.9 lists the estimates for A and E that are unique to preference in addition to the amount of variance shared with total quantity. The heritable component in females was again higher (A_{wine}= 56.4%, A_{beer}=70.3%, A_{DS}= 44.5%) than in males (A_{wine} =21.3%, A_{beer} = 45.2%, A_{DS} = 19.1%). Unlike in the VATSPSUD, the common environmental component could be dropped; therefore, the unique environmental estimates were lower in females (E_{wine} = 43.6%,

E_{beer} =29.7%, E_{DS} = 55.5%) than in males (E_{wine} =78.7%, E_{beer} = 54.8%, E_{DS} = 80.9%). As seen in the VATSPSUD, there was only small to modest overlap in genetic influences on alcoholic beverage preference and the regular consumption measure (2%-28%).

Table 3.10. Model fitting results for regular consumption last week/ABP in the VA30K

Model	Model Comparison	-2LL	$\Delta\chi^2$	Δdf	*p*-value	ΔAIC
Regular Quantity/Wine Preference						
1a. AfCfEf ≠AmCmEm, *ra, rc, re*	-	18475.8	-	-	-	-
2a. AfCfEf =AmCmEm, *ra, rc, re*	2a v. 1a	18486.8	11.0	9	0.27	-7.0
3a. CfEf ≠CmEm, *ra, rc, re*	3a v. 2a	18494.6	7.8	3	0.05	1.8
4a. AfEf ≠AmEm, *ra, rc, re*	4a v. 2a	18499.4	12.6	3	0.01	6.6
5a. AfCfEf ≠AmCmEm	5a v. 2a	18576.2	89.4	3	0.00	83.4
Regular Quantity/Beer Preference						
1b. AfCfEf ≠AmCmEm, *ra, rc, re*	-	19044.4	-	-	-	-
2b. AfCfEf =AmCmEm, *ra, rc, re*	2b v. 1b	19072.3	27.9	9	0.00	9.9
3b. CfEf ≠CmEm, *ra, rc, re*	3b v. 2b	19091.2	18.9	3	0.00	12.9
4b. AfEf ≠AmEm, *ra, rc, re*	4b v. 2b	19081.6	9.3	3	0.03	3.3
5b. AfCfEf ≠AmCmEm	5b v. 2b	19151.5	79.1	3	0.00	73.2
Regular Quantity/Spirits Preference						
1c. AfCfEf ≠AmCmEm, *ra, rc, re*	-	19438.0	-	-	-	-
2c. AfCfEf =AmCmEm, *ra, rc, re*	2c v. 1c	19444.6	6.6	9	0.68	-11.4
3c. CfEf ≠CmEm, *ra, rc, re*	3c v. 2c	19462.2	17.6	3	0.00	11.6
4c. AfEf ≠AmEm, *ra, rc, re*	4c v. 2c	19448.5	3.9	3	0.27	-2.1
5c. AfCfEf ≠AmCmEm	5c v. 2c	19494.5	50.0	3	0.00	44.0

Notes: ABP, alcoholic beverage preference; Af/m, additive genes in females/males; Cf/m, common environment in females/males; Ef/m, unique environment in females/males; LL, log likelihood; $\Delta\chi^2$, change in chi-square from full model; *p*-value associated with change in chi-square; ΔAIC, Akaike's criterion from full model; *ra*, genetic correlation; *rc*, shared environmental correlation; *re*, unique environmental correlation.

The sex differences in parameter estimates seen in the VATSPSUD and FT12 samples were not maintained in the VA30K. Table 3.10 shows that models in which the parameter estimates were equated across the sexes (Model 2a-c) did not fit significantly worse by ΔAIC

and *p*-value than models in which the parameters were estimated separately (Model 1a-c). Therefore, we proceeded with model fitting using sex-equated parameters.

We found that A (Models 3a-c) and C components (Models 4a-c) could not be dropped for any preference, except for the C components in distilled spirits (4c). For comparison purposes across the preferences, we present the estimates for the full models (Table 3.11). The genetic and environmental correlations between last week's total quantity and the preference phenotypes could not be dropped simultaneously (Model 5a-c) across all three preferences. Again, as seen in Table 3.11 and in the two samples, the overlap of genetic factors between the regular consumption measure and preference was small to modest (wine=36%, beer=5%, and DS=1%).

Table 3.11. Parameter estimates for bivariate Cholesky decomposition for ABP and liability shared with regular quantity last week in the VA30K

		Both Sexes		
		Estimate for Preference	With Last Wk Quantity	Unique
Wine Preference	A	12.1	36	64
	C	33.6	1	99
	E	54.3	3	97
Beer Preference	A	39.3	5	95
	C	24.7	6	94
	E	36.0	1	99
Spirits Preference	A	42.5	1	99
	C	6.1	2	98
	E	51.4	4	96

Notes: A, proportion of variance due to additive genetic effects; C, proportion due to common environment; E, proportion due to unique environment

Extended Twin Modeling in VA30K

We found that there were no qualitative or quantitative sex differences when we fitted the full *Cascade* model to the ordinal preference data. As listed in Table 3.12, across wine, beer, and

DS preferences, there was a modest additive genetic component (A_{wine}=21%, A_{beer}=15%, A_{DS}=19%) and little to no non-additive genetic variance (D_{wine}=0%, D_{beer}=3%, D_{DS}=0%). We found no evidence for cultural transmission for any preference and the sibling environment ranged from none to small (S_{wine}=11%, S_{beer}=7%, S_{DS}=0%). The special twin environment was also small to modest (T_{wine}=2%, T_{beer}=18%, T_{DS}=14%), while the unique environment represented the greatest source of variance (E_{wine}=65%, E_{beer}=57%, E_{DS}=67%). Spousal correlations were modest to moderate across all three preferences (wine= 0.32; beer=0.21; spirits=0.29).

Table 3.12. Parameter estimates for univariate extended twin analysis for each ABP in the VA30K

	A	D	F	S	T	E
Wine Preference	0.21	0.00	0.00	0.11	0.02	0.65
Beer Preference	0.15	0.03	0.00	0.07	0.18	0.57
Spirits Preference	0.19	0.00	0.00	0.00	0.14	0.67

Notes: A, additive genetic variance; D, non-additive genetic variance; S, common sibling environmental variance; F, cultural transmission; T, twin environmental variance; E, unique environmental variance.

Discussion

In the first part of this study**,** we sought to examine the relationship between endorsing certain levels of consumption, AUD symptoms and diagnoses, as well as other psychiatric conditions, using ABP in the heaviest year as the determinant**.** All of the studies discussed below, including our own, controlled for several socio-demographic factors shown to be correlated with preference. Our finding that wine preference drinking is associated with the lowest likelihood of AUDs is in agreement with findings of Flensborg-Madsen and colleagues (120). We also found that beer preference drinkers had the greatest odds of high levels and frequency of consumption, which is consistent with previous work examining frequency of drinking and drunkenness (123),

heavy and excessive drinking (122), as well as problem drinking and level of intoxication (125). Further support for our findings comes from research that found that beer accounts for the majority of hazardous alcoholic beverage consumption in the US (81%) (157) and that binge drinkers show a high level of preference for beer (158). Our findings are also reasonable in light of work indicating that beer drinkers are more likely than any other type of drinker to be arrested for drunk driving and involved in alcohol-related accidents (125,159-161). It should be emphasized that while some individuals have the perception that beer is a social drink (123), less harmful than spirits due to its lower alcohol content (123,162), and even safe during pregnancy (163), our results suggest that the risk for AA and most symptoms of AUDs is the same for beer and DS preference drinkers. This, combined with the additional correlates of beer drinking seen in other studies, is reason for reconsideration of beverage-specific public policies in the US, such as lower taxes for beer than DS (162), and for better education regarding the effects of all forms of ethanol.

We also found that DS preference drinkers were at the greatest risk for AD, the most severe form of AUD, and certain symptoms of AUDs that may be considered particularly impairing, including withdrawal, preoccupation, interference with daily activities, and use despite physical/psychological problems. These finding are in line with work by Baltieri and colleagues, who reported that spirit drinkers had a higher severity of AD and craving for alcohol than beer drinkers (164). The only previous study we know of examining depression in relation to ABP found that beer and spirits preference drinkers did not score differently on the Hamilton Depression Rating Scale (164), while our data showed that DS preference drinkers were at greater risk for MD. Compared to beer drinkers, we also found that wine drinkers are at greater risk for MD, which was the only variable we tested that showed this pattern. We are unaware of

any previous work relating ASPD specifically to beverage preference. However, in line with our work, previous research has found that engagement in delinquent behavior in adolescence (126) and assault and homicide rates in adulthood (165) were correlated with spirit and beer drinking but not wine drinking. Our findings on substance use disorders are supported by work indicating that adolescents in Finland who preferred beer or spirits had more exposure to illicit substances than wine or cider preference drinkers (123) and underage youth in the US who preferred these beverages showed greater marijuana use than wine cooler and malt liquor preference drinkers (161). Additionally, beer drinkers had a higher prevalence of illicit drug use than wine drinkers in a Danish adult sample (166).

Along with the results from others studies, the findings from the first part of our report suggest that beverage preference may help define meaningful subtypes of alcohol dependent individuals. A number of typologies currently exist, categorizing dependent individuals using such measures as age at onset, alcohol consumption, comorbidity, and alcohol-related problems (167). Our work suggests that DS preference drinkers are more likely to fall into a chronic/severe subtype, consisting of early onset, severe AD with a high probability of comorbidity, including MD, ASPD, and polysubstance use, while beer preference drinkers are more likely to belong to a subtype with high levels of drinking/drunkenness and legal consequences, but less probability of comorbidity. If wine drinkers develop an AUD, they would likely belong to a later onset subtype, supported by our demographic data of age during the heaviest year, with lower consumption and fewer alcohol-related problems but greater probability of comorbid MD than beer preference drinkers (168). These typologies have implications for treatment efforts, such as that spirits preference drinkers may require more intense intervention and/or a combination of different intervention strategies, while wine and beer preference drinkers may be more responsive to

treatment. Interestingly, Baltieri and co-authors compared the adherence to treatment (topiramate, naltrexone, and placebo) of beer and distilled spirits preference drinkers and found that spirit drinkers were indeed less adherent to drug therapies (164).

There are several plausible explanations for the phenotypic associations between ABP and our outcome variables, including overlapping genetic and environmental liability factors and direct causal effects. The significant co-twin findings for AC measures and ABP indicate that common familial factors account, at least partially, for the correlations. Further evidence for this explanation was provided by the bivariate twin analyses across the VATSPSUD, FT12, and VA30K, which generally showed significant genetic and common environmental correlations between AC measures and preference. Setting is one environmental factor that has been shown to influence alcohol consumption measures as a well as what one chooses to drink. Wine tends to be consumed at home, with meals, and during religious occasions, which provide social constraints on intake, whereas beer and spirits are drunk more frequently at bars, sports games, and parties, where high consumption is tolerated and even preferable (169). A few previous studies have implicated genes that appear to influence both consumption measures and ABP. Ishibashi et al. (170) found that individuals expressing the *ALDH2*2* variant in the alcohol dehydrogenase gene preferred beverages with lower concentrations of ethanol, and this variant is well known to effect consumption measures (171) as well as for risk for AD (44). Additionally, taste sensitivity to the synthetic compound PROP differs based on genetic variability in the bitter taste receptor genes (172), and PROP sensitivity is a good predictor of alcoholic beverage preference (173). Two bitter taste receptor genes, *TAS2R38* and *TASR16,* have been associated with level of alcohol consumption and AD (25,172,174). The majority of co-twin findings for AUDs/symptoms and other psychiatric disorders were not significant, providing less evidence of

common familial influences on these measures and ABP, but there may still be common unique environmental influences.

An additional/alternative explanation for the associations, which is not mutually exclusive from the first, is a causal relationship. As discussed in the introduction, studies suggest that both the concentration and congener differences in specific alcohol beverages may be partially causal for non-psychiatric health outcomes, such as obesity, cancer, and cardiovascular disease (117,119,127). However, there have been few studies on concentration and congeners in terms of their effects on the development of neuropsychiatric outcomes. In a review of experimental studies on the behavioral effects of different types of alcohol, mostly from the 1970's or earlier, Smart (125) notes that subjects generally showed greater psychomotor and intellectual impairment after consumption of spirits than beer. Smart suggested some of these effects may be due to the more rapid rise in blood alcohol concentration after consumption of spirits. A more recent study found that craving, known to be an important factor in relapse, was associated with beer consumption, but not wine or spirits consumption, in males inpatients on a detoxification unit (175). The authors posited that since beer has the highest volume of liquid of the alcoholic beverages, it may alter hormone levels that influence withdrawal craving more than the other beverages. Furthermore, wine and spirits have been shown to increase homocysteine levels, while beer has no effect or perhaps a negative effect (176); increased homocysteine is associated with craving and withdrawal seizures (177). Wilhem et al. also suggest that the hippocampal volume loss that they found to be associated with wine and spirits drinking, but not beer drinking, was due to homocysteine level changes directly caused by these beverages (178). Another study found a significantly increased risk of dementia in beer drinkers and decreased risk in wine drinkers with no effect in DS drinkers (179). Thus, certain alcoholic beverages may

confer neurotoxic effects that result in neuropsychiatric outcomes. Finally, while the main effect of alcohol on hangover appears to be from ethanol, additional effects on hangover severity have been demonstrated depending on congener content of various alcoholic beverages (180,181).

Causal explanations are only logical if the temporal ordering is such that the effect occurs subsequent to the cause. Because we have cross-sectional data, we cannot be certain of the chronological order in which preference and the variables that we treated as outcomes developed. It is possible that individuals establish/change their preference after they develop drinking patterns, AUDs/symptoms, and other psychiatric conditions. However, we believe our assumed temporal ordering is reasonable because longitudinal studies have shown that individuals who prefer wine are less likely to develop into excessive drinkers (122) or become alcohol dependent (120). Additionally, prior literature suggests that preference develops during the teenage years for many individuals (123,161). Other work indicates that preference is relatively stable in several populations (182). In our twin data, both the average age of individuals during the year when AC measures were assessed (27.2 years old, *SD*=9.44) and the average age at onset of the occurrence of two or more symptoms in the same year (22.1 years old, *SD*= 5.79) were past the teenage years.

A final explanation for the associations is residual or uncontrolled confounding. We controlled for several demographic variables that we found to be associated with preference and to which we had access; however, our covariates may not have been adequate. For example, socioeconomic status is known to be a strong correlate of ABP (183), and while we attempted to control for this variable with education level, there may still be residual confounding. Additionally, there are other variables to which we did not have access that the extant literature

has shown to be correlates of preference and drinking patterns/AUDs/other psychiatric conditions, including personality characteristics (155) and coping styles (184).

We are unaware of any studies that have assessed the association between ABP and particular outcomes in the context of a twin study. When a twin's preference was used to assess the odds of reporting particular levels of AC measures, most ORs were significant, suggesting that the association between ABP and consumption levels is at least partly influenced by familial factors. To further explore this relationship, we examined the genetic and environmental influences on regular frequency of consumption and ABP in bivariate twin models in the VATSPSUD and these influences on regular quantity of consumption and ABP in the VA30K and FT12. Generally, we found that total AC measures index genetic liability to the three preferences only modestly. Given that beverage preference may help define certain subtypes of alcohol dependent individuals and examining subtypes are a potential way to reduce genetic heterogeneity, it may be more informative to retain particular beverages as separate measures in association studies instead of combining them into one measure. Genes that influence preference for particular alcoholic beverages may act as modifiers of AUD characteristics (185). For example, being more tolerant of bitter tastes may lead to a preference for spirits, earlier onset of regular drinking, and higher levels of alcohol consumption. Additionally, genes associated with preference may influence the metabolism of particular congeners. Knowing these genes could aid in prediction of individuals at higher risk for negative effects from certain alcoholic beverages, such as red wine headaches (186), and those who are most likely to benefit from positive effects.

In males, the parameter estimates for ABP in the VATSPSUD and FT12 suggest that preference is more influenced by environmental factors. However, in females, we found that

additive genetic factors had a stronger influence. Sex differences in alcohol-related phenotypes are not unusual in the twin literature. In the same twin sample used for the present study, our group found differences between the genders in the heritability for AC measures (9) as well as variability in the sources of genetic and environmental factors contributing to AD (187). Other groups have seen significant sex differences in the heritability of alcohol misuse (188) and AD symptoms (189). Furthermore, a number of studies examining preference for a variety of food items also showed sex differences (146,190,191).

There are several possible explanations for the higher environmental influences on beverage preference in males. Prior research suggests that, in comparison to females, males are more motivated to drink to fulfill conformity motives (fitting in with a peer group) and more influenced by peers in the amount of alcohol they consume because they regard confederate drinking as a challenge (184,192). It may follow that they are more influenced by peers in the type of alcohol that they consume. Landrine and colleagues (193) found that undergraduate subjects attributed beer drinking significantly more often to males, indicating that drinking this beverage is part of gender expectations for men. Work by Corcoran et al. supports this finding (194). The researchers divided males and females into two groups based on their scores on The Fear of Negative Evaluation (FNE) scale, which measures level of apprehension over potential negative social evaluation. The participants sampled four beverages (pineapple juice, diet cola, tonic water, and beer) and were then presented with a series of questions about their preferences. High-FNE males were more likely to select beer than low-FNE men or either group of women. Furthermore, certain sets of genes that affect beverage preference, such as those that encode the bitter taste receptors, may be differentially expressed in the sexes. Interestingly, a higher number of females are classified as supertasters (highest level of bitterness and irritation/astringency)

(195,196), suggesting that females may be more influenced by taste, and thus genetic factors, when selecting their preferred alcoholic beverage.

We are not certain why the VA30K data did not show the same gender discrepancy in parameter estimates as the VATSPSUD and the FT12 samples, but there are several possible explanations. The latter two studies assessed consumption measures and preference over a longer period of time, whereas the VA30K collected information about "last week," which may be more variable, and thus less powerful, as well an atypical week for some individuals. The range of birth years is greater for the VA30K (1915-1971) (148) as compared to the VATSPSUD (1934-1970) (91) and FT12 (1983-1987) (143). A cohort effect on ABP has been seen, which could affect parameter estimates (197). Additionally, participants in the latter two samples were interviewed by clinically trained interviews, whereas subjects in the third sample filled out questionnaires on their own.

In terms of our full *Cascade* model estimates, we saw less evidence for additive genetic influences on all three preferences than the bivariate models in the VA30K and no evidence for cultural transmission. It is not uncommon to see variable parameter estimates based on the number of relatives and little to no evidence of cultural transmission in traits that have been previously assessed in the literature (153,198). The other common environmental estimates indicate that shared environmental variance in preference is more due to special twin environment (beer and spirits) or sibling environment (wine). These influences could be from peers, financial factors, and/or country specific-trends. For example, European countries can be divided into "alcohol belts," based on their production and consumption of wine (Portugal, Spain, Greece, Italy, and France), beer (Germany, Belgium, Denmark, and the United Kingdom) and distilled spirits (the Czech Republic, Russia, Norway, Finland, Sweden, and Poland).

The results of our work must be considered in the context of several limitations. First, we did not ask the subjects directly about their ABP, but rather assumed that their most frequently consumed beverage (VATSPSUD) or the quantity with they consumed particular beverages (FT12 and VA30K) would reflect their preferences. Alcoholic beverage choice may not always reflect preference; however, Straus and Bacon (199) showed that the two traits are highly correlated in college students. While some authors have also utilized "most frequently used beverage" synonymously with ABP (155), others have based preference on a quantity measure. Quantity-based definitions are variable in the literature, including defining preference as the "most consumed beverage" (200), the type of alcohol usually consumed during the past 30 days (161), >35% of alcohol consumed (120), and $\geq$ 50% of total alcohol used (115,122,123,201-203), $\geq$ 2/3 of total consumption (164,204) . It is reassuring to note that definitions based on quantity may not be very different than those based on frequency. Using continuous measures of the quantity and frequency of intake of specific beverages from three datasets to which we have access, including FT12, VA30K, and an AD sample, the correlations between quantity and frequency for wine, beer, and spirits were high, ranging from 0.66-0.86. However, it still would be helpful to have a standard definition of preference throughout the alcohol literature, so that studies using this trait could be more directly comparable.

Second, we did not have specific information on the form of spirits (straight or mixed) or types of beverages within main classes (e.g. red vs. white wine or fortified vs. regular wine). We also did not have access to data on additional alcoholic beverages beyond the three main classes, such as alcopops, cider, and caffeinated alcoholic beverages, which have been shown to have differential associations as compared to wine, beer, and spirits (123,205). Third, it would have been interesting to test for significant differences in the magnitude of genetic and environmental

influences on quantity of consumption of particular alcoholic beverages using multivariate modeling approaches, as has been done with different forms of caffeine, cannabis, and cocaine (146,206). However, this is more challenging with forms of alcoholic beverages in that individuals tend to drink multiple classes of beverages, and so they are more dependent on each other than other forms of drugs of abuse. Finally, it is unclear whether our samples of Virginia- and Finnish-born Caucasians generalizes to other populations. Nevertheless, in the first part of the study in which we attempted to replicate and extend others' work on beverage preference correlates, our findings are broadly similar to the results of studies using Finnish, Danish, Brazilian, and American samples.

CHAPTER 4: Limited Associations of Dopamine System Genes with Alcohol Dependence and Related Traits in the Irish Affected Sib Pair Study of Alcohol Dependence

Adapted from: Hack LM, Kalsi G, Aliev F, Kuo P-H, Prescott CA, Patterson DG, Walsh D, Dick DM, Riley BP, Kendler, KS. Limited Associations of Dopamine System Genes with Alcohol Dependence and Related Traits in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD). *Alcohol: Clinical & Experimental Research.* 2011, 35(2): 376-385.

Abstract

Background: Over 50 years of evidence from research has established that the central dopaminergic reward pathway is likely involved in alcohol dependence (AD). Additional evidence supports a role for dopamine (DA) in other disinhibitory psychopathology, which is often comorbid with AD. Family and twin studies demonstrate that a common genetic component accounts for most of the genetic variance in these traits. Thus, DA related genes represent putative candidates for the genetic risk that underlies not only AD, but also behavioral disinhibition. Many linkage and association studies have examined these relationships with inconsistent results, possibly due to low power, poor marker coverage, and/or an inappropriate correction for multiple testing.

Methods: We conducted an association study on 10 DA related genes (*DRD1-D5, SLC18A2, SLC6A3, DDC, TH, COMT)* using a large, ethnically homogeneous sample with severe AD (N=545) and screened controls (N= 509). We collected genotypes from linkage disequilibrium (LD) tagging SNPs and employed a gene-based method of correction. A total of 135 SNPs were genotyped using the Illumina GoldenGate and Taqman Assays-on-Demand protocols. We tested for association with AD diagnosis in cases and controls and with a variety

of alcohol-related traits (including age-at-onset, initial sensitivity, tolerance, maximum daily drinks, and a withdrawal factor score), disinhibitory symptoms, and a disinhibitory factor score in cases only. We also sought to replicate association results from the Study of Addiction: Genetics and Environment (SAGE) sample between specific DA SNPs and the phenotype sensation seeking (207).

Results: Of the 101 SNPs entered into standard analysis, 6 independent SNPs from 5 dopamine genes were associated with AD or a quantitative alcohol-related phenotype (ARP). 2 SNPs across 2 genes were associated with a disinhibitory symptom count, while 1 SNP in *DRD5* was positive for association with the general disinhibitory factor score. We did not replicate the findings from the SAGE sample in our Irish sample.

Conclusions: Our study provides evidence of modest associations between a small number DA related genes and AD as well as a range of ARPs and measures of behavioral disinhibition. While we did conduct gene-based correction for multiple testing, we did not correct for multiple traits because the traits are correlated and there is a priori evidence of association with each trait. However, false positive findings remain possible, so our results must be interpreted with caution.

Introduction

Owing to its involvement in a broad range of functions, alteration in dopamine (DA) activity appears to play a central role in the etiology and/or treatment of many psychiatric disorders. DA's posited role in alcohol dependence (AD) stems from its involvement in the mesocorticolimbic reward pathway, which spans from the ventral tegmental area to the nucleus accumbens and prefrontal cortex (208). DA was first implicated in mediating the effects of

reward in Olds' and Milner's (209) classic experiments. Subsequent behavioral studies have generated considerable additional evidence to show that dopaminergic transmission in the mesocorticolimbic pathway is essential to reinforcing reward (210).

Of all addictive substances, alcohol has created one of the greatest societal burdens (3). AD is a clinically and etiologically heterogeneous condition that is 50%-60% heritable (7,8). Due to its etiological heterogeneity, considering subtypes of individuals with AD may increase power to detect underlying susceptibility variants. A recent latent class analysis of our sample found that cases could be divided into three classes based on comorbidities: a severe (S) class with the highest probabilities of all comorbidities and high novelty seeking; a depressed (D) class with the highest probability of neuroticism and high probability of depression; and a mild (M) class with the lowest probabilities of all comorbidities (211). These classes are consistent with the idea that at least two independent pathways to the development of AD exist, including negative affect regulation, in which alcohol consumption is a means of relieving negative mood states, and behavioral disinhibition, in which high consumption is part of an overall tendency to behave impulsively and to seek excitement (212). The etiologically independent nature of these pathways is supported by evidence that one common factor is largely responsible for the genetic susceptibility to internalizing disorders, while another common genetic factor explains most of the genetic variation in externalizing disorders, and there is only a small correlation between the two factors (28). These common genetic liabilities may help explain why internalizing phenotypes, such as depression and anxiety/neuroticism, and disinhibitory phenotypes, including drug dependence (DD), antisocial behavior, and attention deficit/hyperactivity disorder (ADHD) are highly comorbid in many samples, including our Irish sample (5,29).

Since the mescorticolimbic pathway may be involved in the rewarding aspects of externalizing behavior, DA genes are reasonable candidates for susceptibility to AD as well as other disinhibitory psychopathology. We considered AD, ADHD, antisocial personality disorder (ASPD), conduct disorder (CD), DD, and novelty seeking (NS) to be part of the externalizing spectrum. Several studies suggest that a disinhibitory personality style, including NS, shares a common genetic influence with disorders in the externalizing spectrum (30,213,214). Childhood/early adolescent studies suggest ADHD share genetic liability with CD (215-220). Additionally, a recent twin study reported shared genetic influences between adolescent ADHD and adult AD beyond those shared with CD (110).

We examined ten DA related genes, including the following: the five receptors, *DRD1-D5*; two transporters, solute carrier family 18 member A2 (*SLC18A2* or vesicular monoamine transporter type 2, *VMAT2*) and solute carrier family 6, member 3 (*SLC6A3* or dopamine active transporter, *DAT* or DAT1); and three enzymes, tyrosine hydroxylase (*TH),* dopa decarboxylase (*DDC),* and catechol-*O*-methyltransferase (*COMT*). If no studies of association between a particular gene and trait are discussed below, we are unaware of any reports (either positive or negative) with rigorous methodology that have examined these associations.

DRD2

DRD2 (11q22-q23) has been examined most thoroughly in relation to AD and disinhibitory phenotypes. The majority of human studies have focused on the TaqIA restriction fragment length polymorphism (rs1800497), which a meta-analysis estimated to be associated with AD at a modest odds ratio (OR) of 1.31 (57). In 2004, Neville et al. (58)found this SNP to be within the coding region of the neighboring ankyrin repeat and kinase domain containing 1

(*ANKK1*) gene; therefore, the most parsimonious explanation is that phenotypic associations are due to this nonsynonymous coding change in *ANKK1*. However, this SNP could be tagging a polymorphism in *DRD2*, which contains several SNPs in modest LD ($r^2 \approx 0.6$) with rs1800497. Furthermore, Dick et al. (59) found weak associations between variants in *DRD2* and AD. Studies examining association with *DRD2* and aspects of heroin dependence have generally been positive (57). The meta-analysis of Gizer and colleagues (221) identified no association of *DRD2* with ADHD.

Other DA receptors

Rigorous studies of *DRD1* (5q35.1) and AD have typically estimated positive associations with modest effect sizes (60,61), while reports examining this phenotype and *DRD3* (3q13.3) have been negative (57). A recent review supports an association between the 48-bp variable number of tandem repeats (VNTR) in *DRD4* (11p15.5) and an intermediate phenotype termed urge for addictive substances, which refers to craving for substances of abuse (62). We know of no rigorous reports investigating associations between *DRD5* (4p16.1) and AD or related traits.

Le Foll and colleagues' 2009 (57) review notes that certain variants in *DRD3-D5* confer an increased risk of heroin dependence, while research on psychostimulant dependence has either been negative or inconclusive. A recent meta-analysis examining ADHD reported significant associations with variants in both *DRD4* and *DRD5*, whereas there was no association with *DRD3* (221). Additionally, Kim and colleagues (61) identified variants in *DRD1* that increase scores for the disinhibitory personality trait novelty seeking (NS) in alcohol dependent subjects. Of all DA genes, *DRD4* has been the best studied for its role in NS. Munafò et al.'s

2008 (222) meta-analyses identified association with C521T (rs1800955) but not with *DRD4's* 48-bp VNTR.

Transporters

SLC18A2 shuttles cytosolic monoamines into synaptic vesicles. Schwab et al. (223) reported an association with variation in *SLC18A2* (10q25) and AD, but we are unaware of any reports of association with other disinhibitory psychopathology. SLC6A3 terminates DA signaling by removing this neurotransmitter from synaptic clefts. Some research has indicated that striatal SLC6A3 density and availability is reduced in alcohol dependent subjects (224). Van der Zwaluw's 2009 (225) review notes that many investigators have identified association between *SLC6A3*'s (5p15.3) best studied 40-bp VNTR and alcohol-withdrawal symptoms but typically not with AD. Generally, association studies of other drug use phenotypes and *SLC6A3* have been negative (226,227), although Guindalini and colleagues (228) did find that alleles in a 30-bp VNTR increased risk of cocaine abuse. Additionally, there have been mixed results of association with *SLC6A3's* functional 40-bp VNTR in relation to antisocial behavior in adolescence (229-233). The meta-analysis of Gizer and coauthors (221) found association of variants in *SLC6A3* with ADHD.

Enzymes

TH is the rate limiting enzyme in DA synthesis. Dahmen et al. (234) showed an increased frequency of the Val allele (Val81Met polymorphism) of *TH* (11p15) in patients with early onset AD. Association studies of ADHD have been negative (235), and no rigorous reports have examined any of the other disinhibitory traits. The final enzyme in the synthesis pathway, DDC,

converts L-DOPA to DA. No published studies have reported an association with *DDC* (7p12.2) and AD or any disinhibitory phenotypes that we studied, including ASPD, ADHD, CD, DD, or NS.

COMT is a degradatory enzyme for catecholamines. The most well-researched polymorphism in *COMT* (22q11.21) is the common G>A transition (rs4680), which results in a valine to methionine substitution (Val158Met) and a decrease in enzyme activity by 3- to 4-fold (236). Many studies have investigated the association between the low activity allele (Met) and AD with inconsistent results (237). Investigators have reported associations between this SNP and methamphetamine abuse (226) and NS (238,239), while Cheuk et al.'s (240) meta-analysis estimated no association with ADHD.

Study Goals

The primary goal of the present study was to test for association between AD and several ARPs, such as initial sensitivity, and SNPs in 10 DA genes in a large, homogeneous sample. Because twin studies have found that a single genetic factor is largely responsible for the genetic susceptibility to AD and several other disinhibitory phenotypes and traits, a second goal of the study was to test for association of these SNPs with relevant symptoms of disinhibitory disorders as well as a disinhibitory factor score. Our final goal was to attempt replication of work by Derringer and colleagues in which the authors tested for association of 273 LD-tagging SNPs across 8 DA genes with the phenotype sensation seeking (207). Their sample included 635 participants from the SAGE sample of which 65.2% met criteria for lifetime AD using DSM-IV criteria (6). The authors found 12 of these SNPs to be individually associated with sensation seeking. Furthermore, they showed that a linear regression model that included the SNPs fit better and explained 3.9% more of the variance in

sensation seeking than one in which covariates alone were included. While we did not measure sensation seeking in our subjects, previous work has found a moderate and significant correlation between this phenotype and one of our phenotypes, novelty seeking (241). Therefore, we assessed association between NS and specific DA SNPs both individually and in aggregate in our replication attempt. To our knowledge, we are the first group to report results on an association analysis of both quantitative ARPs and symptoms of disinhibitory disorders within alcohol dependent cases for a large group of DA related genes.

Methods

Subjects and phenotype measurement

Participants in the Irish Affected Sibpair Study of Alcohol Dependence (IASPSAD) were recruited in Ireland and Northern Ireland between 1998 and 2002. Further details of the study design, sample ascertainment, and clinical characteristics of this sample are described elsewhere (242). In brief, ascertainment of probands was mainly conducted in community alcoholism treatment facilities and public and private hospitals. Probands were eligible for study inclusion if they met the current DSM-IV criteria for AD and if all four grandparents had been born in Ireland, Northern Ireland, Scotland, Wales, or England. After a prospective family was identified through probands, parents and potentially affected siblings whom the probands provided permission to contact were recruited.

Probands, siblings, and parents were interviewed by clinically-trained research interviewers, most of whom had extensive clinical experience with alcoholism. The assessment included demographic characteristics, lifetime history of AD and other comorbid conditions, ARPs, personality features, and clinical records. All participants provided informed consent. Controls were recruited in Northern Ireland from volunteers donating at the Northern Ireland Blood Transfusion Service and in the Republic from the Garda Siochana (the national police force) and the Forsa

Cosanta Aituil (the army reserve). Controls were screened and their samples excluded if they reported a history of heavy drinking or problem alcohol use. In the present case-control study design, we included 545 independent AD cases and 509 controls with an ample yield of high quality DNA for genotyping. The selection of cases was random with respect to AD severity and comorbid phenotypes.

In addition to the binary diagnosis of AD, we chose to examine several quantitative ARPs because examining such traits may provide more power than analysis of dichotomous phenotypes (20); prior research in this sample detected linkage and association signals with some of these traits but not with AD (e.g.(243)). We assessed age-at-onset of AD (ONSET), subjective response to ethanol, maximum drinks in 24 hours (MAX24), and a factor score of withdrawal symptoms (WDSFS). ONSET was defined as the age at which the first criterion for DSM-IV AD was satisfied. Subjective response to ethanol was assessed using the self-rating of the effects of ethanol (SRE,(244) to form two scores, initial sensitivity (ISENS) and tolerance/maximum drinking (TOLMX). The SRE inquires about how many drinks were needed for a subject to experience effects from alcohol consumption at different stages of use. ISENS is based on "the first 5 times you ever drank" and items contributing to TOLMX concern the "period when you drank the most". The score of each measure was computed by summing the number of drinks required to produce an effect and dividing by the number of effects endorsed. The SRE has been shown to have good internal consistency and test–retest reliability, to successfully identify people who had low response to alcohol in a laboratory challenge test, and to be associated with AD diagnosis in several populations. Due to non-normal distributions of the regression residuals, we log transformed values for ISENS and TOLMX. MAX24 refers to the largest number of drinks an individual reported ever having consumed in 24 hours. The withdrawal severity factor score was based on ten symptoms in the Semi-Structured Assessment of

the Genetics of Alcoholism (SSAGA, version 11, (142)) interview (such as hands trembling, feeling anxious following cessation or reduction of drinking). To account for the possible non-equal contribution of each symptom to withdrawal severity, a factor analysis was conducted (for details see (243)). A factor score of withdrawal severity for each individual was derived based on the item loadings on one major factor, which accounted for 70% of the variance in these symptoms based on the entire IASPSAD sample.

We also tested the DA genes for association with scores for disinhibitory disorders. Symptom counts for lifetime DSM-IV alcohol dependence (ADsx, range 3-7) were assessed using the SSAGA, modified to reduce assessment time. Counts for number of illicit drugs fulfilling criteria for substance dependence (DDsx, range 0-7), conduct disorder symptoms (CDsx, range 0-14), and antisocial personality disorder symptoms (ASPDsx, range 0-9) were collected using adapted versions of the Structured Clinical Interview for DSM-IV Disorders (SCID, (92)). Drugs assessed as part of illicit substance use included cannabis, sedatives, stimulants, cocaine, opiates, hallucinogens, and other drugs (e.g. steroids, nitrous oxide). All drugs were considered illicit in this context because the subjects were asked only about nonmedical use (e.g. use without a prescription, use in greater amounts/more often than was prescribed, or use not in the intended manner). Other measures include retrospective reports of childhood attention deficit/hyperactivity disorder symptoms (ADHDsx) using items from Wender's Childhood Problem Behavior Checklist (245) and novelty seeking scores (NS) using the 18-item version from Cloninger's Tridimensional Personality Questionnaire (246). Scores for ADHDsx and NS were rescaled such that their range is 0-1. Finally, we tested for association with a factor score based on the item loadings on one major factor for all disinhibitory phenotypes. We modeled this analysis off association analyses previously conducted in the Collaborative Study on the Genetics of Alcoholism and the Virginia Adult Twin Study for Psychiatric and Behavioral Genetics

(31,247). Due to non-normal distributions, we log transformed ADsx, ASPDsx, CDsx, and DDsx for the factor score analysis.

Tag SNP selection and genotyping

The majority of genotyping was conducted in Dr. David Goldman's lab at NIAAA on the Addiction Array (248) using the Ilumina GoldenGate method. For details of study design, see reference (248). In instances where additional SNPs had to be genotyped in order to complete tagging in our sample, we selected LD-tagging SNPs (tSNPs) with Tagger (249) as implemented in Haploview 3.2 (250) using the default criteria of $r^2 \geq 0.8$ and minor allele frequency (MAF) ≥ 0.2. Because SNPs genotyped using the Ilumina GoldenGate platform were chosen based on being African haplotype tagging, some SNPs have a MAF< 0.2 in our Caucasian sample. However, no SNP has an MAF< 0.01, which was our threshold for eliminating SNPs in the overall sample. For genes displaying several isoforms, the longest isoform was chosen for tag selection but in order to limit genotyping load and cost, 5' and 3' regions of the genes and ESTs were not directly tagged.

tSNPs were genotyped in-house as monoplex reactions using Taqman Assays-on-Demand (Applied Biosystems, Foster City, CA). To ensure uniformity and accuracy, all reaction steps were performed using the Eppendorf 5075 automated liquid handling platform. Stringent evaluation of initial data is important to avoid artifactual 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. Individual DNA samples with 20% or more missing genotypes across the entire study were also excluded. Individual SNPs were excluded if they showed deviation from Hardy-Weinberg equilibrium ($p > 0.001$) in the overall sample and controls alone, had an MAF<0.01 or had a low genotyping call rate (<80%). The present study reports on the results

from genes included in the DA functional domain (248), including *DRD1-5*, *SLC18A2*, *SLC6A3*, *DDC*, *TH*, and *COMT*. Several coding SNPs were genotyped, such as rs155417 and rs5326 in *DRD1*, rs6279 in *DRD2*, rs6347 in *SLC6A3*, rs11575542 and rs11575377 in *DDC*, and the well studied nonsynonymous SNP rs4680 in *COMT*. The well known TaqIA polymorphism in *DRD2* was not included because, as noted above, it is actually located in a different gene.

Statistical methods

Single marker analyses were implemented in PLINK 1.07 (251) using logistic regression for the binary trait of AD and linear regression for the quantitative traits in cases only to calculate effect size (either odds ratio or regression coefficient) and significance level. We used sex as a covariate in the logistic regressions and both age and sex as covariates in the linear regressions. To address the possibility of type I error due to multiple testing of several SNPs within individual genes, we permuted each *p*-value 10,000 times using the gene-based set test in PLINK and only reported the empirical *P*-value here if it was significant after this correction. We reasoned that gene-based correction was sufficiently conservative because all selected genes have a priori evidence of association with AD and/or related phenotypes. We used the set-based test in PLINK for multiple test correction because this method allows for identification of independent SNPs determined by a selected threshold. We changed the default threshold for LD from r^2=0.5 to r^2=0.8 because our LD-tagging SNPs were selected based on $r^2 \geq 0.8$. Additionally, we did not correct for multiple phenotypes because (1) they reflect reasonable a priori hypotheses, and (2) they are, in some cases, substantially inter-correlated, making it difficult to implement any simple multiple test correction.

The goal of principal component analysis is to reduce the number of observed correlated variables in a data set by uncovering latent dimensions (principal components, factors). We used the FACTOR procedure in SAS version 9.2 (93) to determine the structure of our phenotypes: ADsx, ASPDsx, ADHDsx, CDsx, DDsx, and NS. This procedure offers a number of methods for extracting factors, including principal component and principal factor analysis. The former method decomposes the total variance, or both the variance shared between variables and that which is unique to individual variables, while the latter decomposes only the variance shared between the variables. We chose principal component analysis for this study. Several options for rotation of factors are available to produce more interpretable results; we chose an orthogonal rotation procedure called VARIMAX (variance maximizing) in which the components are forced to be uncorrelated and each variable is associated with only one component. We then calculated component scores for each individual with nonmissing data for all disinhibitory phenotypes and tested for association with these scores.

We utilized the statistical program R (252) to assess the fits of two linear regression models for predicting variation in novelty seeking. In the first (restricted) model, we included the covariates age and sex only. In the second (unrestricted) model, we added the effects of 5 DA SNPs that were identical or proxies for those in the SAGE study to the base model. These SNPs are located in *DDC* and *COMT*. We then compared the amount of variance in novelty seeking explained and the fit of each model using an *F* test and its associated *p*-value.

Power estimates for this study were calculated using QUANTO 1.2.4 (http://hydra.usc.edu/gxe) for the dichotomous AD outcome (Table 4.1) and the continuous traits (Table 4.2) using a two-sided t-test with a significance level of 0.05 and a range of MAF from 0.05-0.4. We assumed an additive mode of inheritance. Effect sizes are listed as OR from 1.1 to

Table 4.1. Power for dichotomous AD outcome

MAF	OR	Power
0.05	1.1	0.0765
	1.2	0.1526
	1.3	0.2711
	1.4	0.4163
	1.5	0.5663
0.1	1.1	0.1006
	1.2	0.2459
	1.3	0.4552
	1.4	0.6665
	1.5	0.8271
0.2	1.1	0.1406
	1.2	0.3908
	1.3	0.6836
	1.4	0.8819
	1.5	0.9676
0.3	1.1	0.1690
	1.2	0.4817
	1.3	0.7895
	1.4	0.9447
	1.5	0.9902
0.4	1.1	0.1855
	1.2	0.5288
	1.3	0.8331
	1.4	0.9635
	1.5	0.9948

Table 4.2. Power for continuous outcomes

MAF	Variance	Power
0.05	0.01	0.6418
	0.02	0.9089
	0.03	0.9814
	0.04	0.9968
	0.05	0.9995
0.1	0.01	0.6418
	0.02	0.9089
	0.03	0.9814
	0.04	0.9968
	0.05	0.9995
0.2	0.01	0.6418
	0.02	0.9089
	0.03	0.9814
	0.04	0.9968
	0.05	0.9995
0.3	0.01	0.6418
	0.02	0.9089
	0.03	0.9814
	0.04	0.9968
	0.05	0.9995
0.4	0.01	0.6418
	0.02	0.9089
	0.03	0.9814
	0.04	0.9968
	0.05	0.9995

1.5 for AD and variation in the traits from 1% to 5% for continuous outcomes. The power for the dichotomous outcome was based on a lifetime population risk for AD of 12.5% (5)

Results

Missingness

Genotyping was completed for 135 SNPs in the DA related genes, but 10 SNPs were excluded due to a low genotyping rate (<80%) and 24 due to a MAF<0.01. The average genotyping call rate for the remaining 101 SNPs was 98.7% (90.6%-100%). All remaining SNPs were in Hardy Weinberg equilibrium (regular cut off *p*-value of 0.001). Genotyping error rate was estimated using duplicates at 0.6%. Among the 1054 genotyped individuals, 26 (8 cases and 18 controls) were excluded due to ≥20% missing genotypes, leaving a total of 1028 individuals. 592 individuals were missing no genotypes, 374 were missing 1-5, 40 were missing 6-10, 15 were missing 11-15, and 7 were missing 16-20. After QC measures were completed, the following number of

SNPs from each gene were entered in standard analysis: 4 SNPs in *DRD1*, 15 in *DRD2*, 13 in *DRD3*, 4 in *DRD4*, 7 in *DRD5*, 11 in *SLC18A2*, 12 in *SLC6A3*, 22 in *DDC*, 3 in *TH*, and 10 in *COMT*.

Principal Component Analysis

Table 4.3 lists the means and standard deviations for the disinhibitory symptom counts as well as the Pearson correlations among counts included in the factor analysis. To determine the number of factors to retain, one can use several guidelines, including the Kaiser (eigenvalue) criterion and visual inspection of the data using a scree plot. Eigenvalues provide the proportion of variance in the total sample accounted for by each component when divided by the number of variables entered into the analysis. In the first method, one chooses to retain components with eigenvalues greater than 1.0. A scree plot is a graph of the components on the x-axis and their corresponding eigenvalues on the y-axis. One finds the place in the plot where the smooth decrease levels off and retains factors to the left of this point. While the Kaiser criterion may sometimes over-extract factors and visual inspection

Table 4.3. Means and standard deviations for disinhibitory symptom counts and Pearson correlations between counts

	ADsx (6.39, 1.03)	ADHDsx (0.50, 0.20)	ASPDsx (2.02, 2.75)	CDsx (2.86, 2.99)	DDsx (.82, 1.4)	NS (.58,0.22)
ADsx	1.00					
ADHDsx	0.18	1.00				
ASPDsx	0.20	0.36	1.00			
CDsx	0.22	0.45	0.78	1.00		
DDsx	0.13	0.19	0.34	0.38	1.00	
NS	0.18	0.31	0.18	0.22	0.19	1.00

Notes: ADsx, alcohol dependence symptoms; ASPDsx, antisocial personality disorder symptoms; ADHDsx, attention deficit/hyperactivity symptoms; CDsx, conduct disorder symptoms; DDsx, drug dependence symptoms; NS, novelty seeking scores. Means and standard deviations are listed in parentheses. ADHDsx and NS scores were rescaled such that their range is 0-1.

using the scree plot may retain too few, both perform relatively well in situations like ours where there are few factors and many subjects (253).

By both a traditional eigenvalue criterion and a scree plot, only one factor was evident with an eigenvalue of 2.55 that accounted for 42.5% of the variance. The eigenvalue difference between the first and second factors was 1.56 and all other factors comprised less than 17% of the variance. The component loadings for each of the symptom counts were as follows: ADsx, 0.41; ADHDsx, 0.65; ASPDsx, 0.82; CDsx, 0.87, DDsx, 0.56, and novelty seeking scores, 0.47. These loadings represent the correlation coefficients between the factor and each variable and can be squared to determine the amount of variance in the dependent variables explained by the factor. The comparatively low factor loading for ADsx may be explained by the restricted range of alcohol dependence symptoms. Subjects could only have within the range of 3 to 7 symptoms and over 80% had $\geq$ 6 symptoms.

Single marker association

In Table 4.4, marker information, empirical *P*-values, and effect sizes are provided for SNPs (1 in *DRD4* and 2 in *SLC6A3)* were associated with AD, 2 SNPs (in *DRD5* and *TH)* with WDSFS, and 1 in *DRD3* with ISENS. Table 4.5 presents the marker information, empirical *P*-values, and effect sizes only for SNPs that were significant after permutation with the disinhibitory symptoms and the disinhibitory factor score. See Table A.2, which contains nominal *p*-values, for a complete listing of all SNPs that underwent analysis for disinhibitory phenotypes. 2 independent SNPs across 2 genes were associated with one or more disinhibitory symptom counts after permutation testing, including 1 SNP in *SLC6A3* with ADsx and 1 in *DDC* with DDsx. Additionally, 1 SNP in *DRD5* was significantly associated with for the factor score.

Table 4.4. Marker information, empirical *P*-values, and effect sizes for SNPs significantly associated with AD or ARPs after permutation testing

GENE & CHR	SNP	BP	LOCATION	CALL RATE	MAF	AD (N=1028)	ONSET (N=436)	ISENS (N=428)	TOLMX (N=427)	MAX24 (N=436)	WDSFS (N=436)
DRD3 (Chr 3)	rs2654754	115338486	intron 5	99.8	.025	--	--	**0.026** (0.24)	--	--	---
DRD4 (Chr 11)	rs12280580	616220	5' near gene	92.7	.354	**0.035** (1.28)	--	--	--	--	--
DRD5 (Chr 4)	rs7655090	9374973	3' near gene	98.1	.049	--	--	--	--	--	**0.0037** (-0.37)
SLC6A3 (Chr 5)	rs27048	1465645	intron 8	97.1	.451	**0.025** (0.68)	--	--	--	--	--
	rs10052016	1481111	intron 4	93.9	.412	**0.025** (1.33)					
TH (Chr 11)	rs11564717	2143465	intron 12	99.9	.408	--	--	--	--	-	**0.025** (0.59)

Notes: AD, alcohol dependence; ONSET, age-at-onset of AD; ISENS, initial sensitivity; TOLMX, tolerance/maximum drinking; MAX24, maximum drinks in 24 hours; WDSFS, withdrawal severity factor score. Empirical *P*-values are in bold. Effect sizes are listed in parentheses as odds ratios for AD and as regression coefficients for quantitative traits. An odds ratio >1 or a positive regression coefficient indicates that the minor allele increases risk of the phenotype. ISENS and TOLMX were log transformed.

Table 4.5. Marker information, empirical *P*-values, and effects sizes for SNPs significantly associated with disinhibitory symptom counts after permutation testing

GENE & CHR	SNP	bp	LOCATION	CALL RATE	MAF	ADsx (N=436)	ADHDsx (N=426)	ASPDsx (N=436)	CDsx (N=436)	DDsx (N=432)	NS (N=429)	Factor Score (N=426)
DRD5 (Chr 4)	rs7655090	9374973	3' near gene	98.1	.049	--	--	--	**--**	--	--	**0.033** (-0.34)
SLC6A3 (Chr 5)	rs6350	1496199	exon 2	99.9	.076	**0.012** (-0.36)	--	--	--	--	--	--
DDC (Chr 7)	rs11575542	50305196	exon 14	99.7	.015	--	--	--	--	**0.014** (0.30)	--	--

Notes: ADsx, alcohol dependence symptoms; ASPDsx, antisocial personality disorder symptoms; ADHDsx, attention deficit/hyperactivity symptoms; CDsx, conduct disorder symptoms; DDsx, drug dependence symptoms; NS, novelty seeking scores. Empirical *P*-values are in bold. Effect sizes are listed in parentheses as regression coefficients. A positive regression coefficient indicates that the minor allele increases risk of the phenotype.

Table 4.6 shows the marker information, nominal p-values, and effect sizes for SNPs that were part of our replication effort. None of these SNPs were individually significant. Table 4.7 shows the characteristics and comparison of model fit for the two models fitted to the data. The unrestricted model did not fit significantly better (F=0.181, *p*-value=0.97) than the restricted one and the additional variance explained by the SNPs was only 0.2%, indicating that these SNPs have no explanatory power for variation in novelty seeking in our Irish sample.

Table 4.6. Marker information, nominal *p*-values, and effect sizes for replication attempt

GENE & CHR	SNP	BP	LOCATION	CALL RATE	EFFECT SIZE (β)	*p*
DDC (Chr 7)	rs11575542	50305196	exon 14	99.7	0.01	0.86
DDC (Chr 7)	rs921451	50397494	intron 1	94.2	0.00	0.83
DDC (Chr 7)	rs6969081	50398714	intron 1	99.7	0.00	0.83
DDC (Chr 7)	rs3829897	50403973	intron 1	99.7	0.00	0.93
COMT (Chr 22)	rs933271	18305961	intron 1	99.6	0.01	0.69

Notes: Effect sizes are regression coefficients. A positive regression coefficient indicates that the minor allele increases risk of the phenotype.

Table 4.7. Comparison of models predicting variation in novelty seeking

MODEL	# of SNPs	R^2	F	Δdf	*p*
Covariates alone	0	8.6%	--	--	--
Covariates plus SNPs	5	8.8%	0.181	5	0.97

Notes: R^2, amount of variance explained by the model; F, F-statistic; df, degrees of freedom.

Discussion

For greater than a half century, dopaminergic dysregulation has been implicated in AD and other disinhibitory psychopathology. Molecular genetics studies over the past 20 years have attempted to demonstrate associations with DA genes and disinhibitory phenotypes, producing

an inconsistent and controversial literature. Meta-analyses, reviews, and reports with rigorous methodology suggest that variation in DA genes does contribute to susceptibility to disinhibitory traits, although not to the extent and effect size originally hypothesized.

We attempted to address some of the problems that have riddled candidate gene studies by using a relatively large, ethnically homogeneous sample with severe AD. We genotyped a sufficient number of tSNPs to cover most of the variation within 10 DA related genes, including *DRD1-D5, SLC18A2, SLC6A3, DDC, TH,* and *COMT*. Our study not only tested for association with the categorical diagnosis of AD, but also with quantitative ARPs, which give more power than dichotomous traits and provide additional clinical information beyond a binary phenotype. Within alcohol dependent cases only, we tested for association with disinhibitory psychopathology, including symptoms for alcohol dependence, antisocial personality disorder, attention deficit/hyperactivity disorder, conduct disorder, drug dependence, and scores for novelty seeking. Additionally, we assessed for association with a disinhibitory factor score. Finally, we only reported findings that were significant after set-based permutation, which limits the possibility that any of our results are false positives.

Overall, we found evidence for association with modest effect sizes between a small number of DA related genes and alcohol dependence, ARPs, and disinhibitory phenotypes. The minor allele frequencies for several of the positive SNPs are low. The limited number of positive signals suggests that these 10 DA related genes play a minor role in susceptibility to alcohol dependence and related disinhibitory psychopathology, which is consistent with previous meta-analyses, reviews, and reports with rigorous methodology. Furthermore, we did not find evidence that the DA SNPs associated with sensation seeking in the Derringer et al. study (207) were significantly associated with novelty seeking either individually or in aggregate in our sample. In

addition to the associations in the first study being false positives or ours being false negatives, there are several other possible explanations for the lack of agreement. One explanation is that while the sensation seeking and novelty seeking are phenotypically correlated, they are not the same phenotype; therefore, it is reasonable to hypothesize that some SNPs that explain variation in one phenotype would not explain variation in the other. Second, SAGE consists of individuals from the US, whereas our sample comes from Ireland; there may be different polymorphisms that contribute to variation in sensation/novelty seeking in each of these ethnicities. Finally, 65.2% of the cases in the Derringer paper were alcohol dependent, whereas all of our subjects were. The lack of replication may be explained by the differing proportion of cases in the two samples.

Receptors

In agreement with reports from several other groups (254-256), we did not find significant association after permutation between the well-studied *DRD3 BaI*I polymorphism (rs6280) and AD. However, we did show that another SNP (rs2654754, p=0.026, β=0.24) in *DRD3* is associated with the quantitative trait of initial sensitivity. Perhaps we obtained these findings when other researchers did not because our sample size is larger, we captured most of the variation in *DRD3* with our 11 tSNPs, and we assessed quantitative ARPs in addition to the dichotomous phenotype. However, the chance that this is a false positive is greater in light of the fact that the MAF=2.3%.

Moreover, we identified an association with 1 SNP in *DRD4* (rs12280580, p=0.011, OR=1.28) and AD. While no other groups have reported associations with this particular SNP and AD or disinhibitory psychopathology, several meta-analyses have noted associations with

variants in *DRD4* and urge for addictive substances (62), ADHD (221), and NS (222). Furthermore, we identified associations with the same SNP in *DRD5* and two phenotypes, including withdrawal (rs7655090, p=0.0037, β=-0.37) and the factor score (rs7655090, p=0.0033, β=-0.34). The fact that *DRD5* is associated with an alcohol-related trait and the factor score suggests it may contribute to AD through the broader disinhibitory spectrum. However, since only one SNP is associated with the factor score and this polymorphism has a low MAF (4.9%), this finding might represent is a false positive.

Perhaps we did not identify strong evidence for association of DA genes with the general disinhibitory factor because our design assessed disinhibitory phenotypes in subjects with alcohol dependence. Another explanation is that variation in dopamine genes may contribute more to risk for specific disorders than the liability to a general disinhibitory spectrum of disorders.

Transporters

Two SNPs in *SLC6A3* (rs27048, p=0.025, β=0.68; rs10052016, p=0.25, β=1.33) were associated with the dichotomous phenotype of alcohol dependence. Within the same gene, we also identified an association between the common allele of another SNP (rs6350, p=0.012, β=-0.36) and alcohol dependence symptoms. Lind and colleagues (224) found association of the same allele in rs6350 with problem drinking in a Finnish population.

Enzymes

We identified association of one SNP in *TH* (rs11564717) with the withdrawal factor score (p=0.025, β=0.59). However, we must note that this SNP has a MAF of only 1.5%. We did

not show association of the Val81Met (rs6356) variant with any ARPs, although Dahmen and colleagues (234) identified association of this polymorphism with early onset AD.

Furthermore, we did not detect any signal with *COMT*'s well-studied functional polymorphism rs4680, which has been associated with a number of disinhibitory phenotypes, including AD (237), methamphetamine abuse (226), and NS (238,239). One explanation for this may be that none of the potential risk alleles were found in our population. It is noteworthy that the SNP in *DDC* (rs11575542, p=0.014, β=0.30) that we identified as associated with drug dependence symptoms is a missense coding polymorphism that results in a substitution from Arg to Gln; however, the MAF of this SNP in our sample is only 2.5%, which increases the likelihood that it is a false positive.

Although, in many ways, we improved the design of previous candidate gene studies, our report still has limitations. First, as noted previously, a proper correction for multiple testing has been problematic in these studies. Using too liberal an approach will maximize power but is likely to lead to false positives. A correction method that is too conservative will decrease power to detect true results. We attempted to strike a balance between the two approaches by using gene-based correction. However, we did not correct for testing multiple phenotypes because there is a priori evidence of association of at least some of the 10 DA related genes with each of the phenotypes examined. Furthermore, many of these phenotypes are highly inter-correlated, making an appropriate correction problematic. It can be argued that our approach is still too liberal. The possibility that some proportion of our findings represents false positives is plausible. Therefore, our findings should be considered tentative pending the outcome of attempted replications. Second, quantitative ARPs and disinhibitory symptoms were measured only among cases but not controls, so the values for these traits are not representative of the full

variation in the population. Thus, the most meaningful replication of our study would be in population sample. Third, although the LD patterns of the DA genes are compatible with the Hapmap CEPH population data, it remains possible that we lack complete coverage of common variation in our Irish sample. Additionally, the impact of rare functional polymorphisms was not assessed. Finally, we did not include all genes that affect dopaminergic tone, such *DBH* (dopamine beta hydroxylase) and the *MAO* (monoamine oxidase) gene.

CHAPTER 5: Preliminary Genome-wide Association Studies of Alcohol Dependence, Symptom Count, and Externalizing Phenotype in an Irish Family Sample

Note: These results are based on preliminary analyses and will likely be updated in the final, published manuscripts.

Abstract

Background: The powerful, systematic, and unbiased GWAS has been successful in identifying replicated susceptibility variants for numerous complex diseases. Compared to several other psychiatric disorders, such as schizophrenia and autism, the total sample size for genome-wide studies of alcohol dependence (AD) is more modest with individual samples in the range of 500-2500 individuals. We report here results from a comparably sized sample (total N=2465) consisting of related cases and unrelated, population controls.

Methods: The related cases are from the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) and were diagnosed using lifetime DSM-IV criteria. Genotyping was conducted using the Affymetrix v6.0 array by three separate genotyping core facilities. Because artifacts are a known issue when combining samples genotyped at multiple sites, genotypes were called using the algorithm BEAGLECALL, which considers both allele signal intensities and LD information. After quality control filtering, the sample contained 710 cases, 1755 controls, and 676,736 SNPs for analysis. The within-site and cross-site duplicate error rates were 0.0043% and 0.0054%, respectively. After imputation using the March 2012 1000 Genomes data and post-imputation filtering, we had a total of 8.2 million SNPs. We analyzed AD, AD symptom count, and an externalizing component score. The binary trait analysis was conducted using MQLS (modified quasi-likelihood score test), while the quantitative trait analyses were run in

ProbABEL; both programs use relationship matrices to correct for the non-independence of siblings. We attempted replication of the binary diagnosis in the European American sample (N=1399) from COGA, which we also imputed to the latest 1000 Genomes panel.

Results: For AD, there were 12 non-independent genome-wide significant signals in the novel risk gene, *COL6A3*. Other top signals fell within additional novel candidate genes, including those involved in neurodevelopmental disorders and cancer. There was little evidence for replication with the COGA sample, but a joint meta-analysis produced additional novel candidates. Likely due to the limited variation of ADsx in our severely affected sample, results from this trait yielded a deflated QQ plot, but there was one intergenic genome-wide significant signal. Finally, we present preliminary results from the externalizing component score analysis.

Conclusions: We have identified some potentially interesting loci associated with AD, AD symptom count, and a composite externalizing phenotype that should be followed-up in functional studies. Ongoing work includes replication of ADsx and the externalizing component score analyses in COGA as well as replication of all three traits in a clinically ascertained German sample.

Introduction

Alcohol dependence (AD), as defined by the Diagnostic and Statistical Manual of Mental Disorders, edition 4 (DSM-IV), is common, relapsing, and under-treated (5,6). Some progress has been made in clarifying the etiology of this serious personal and public health problem, but a more complete understanding is needed to develop more effective prevention and intervention strategies. One approach to improving our understanding of the etiology of AD is identification of susceptibility variants. The powerful, systematic, and unbiased genome wide association study

has been successful in identifying replicated risk variants for several common, complex disorders. Over the last several years, multiple GWAS of AD, AD symptomatology, and consumption measures in both clinical and community samples have been published (46,63-66,68,69). However, there have been fewer genome-wide significant findings and less robust replication of top results than in other psychiatric disorders, such as schizophrenia and nicotine dependence (39,257). One reason is likely that sample sizes for AD studies individually have not reached the level of other psychiatric disorders. Another reason may be that even though most studies define cases using lifetime DSM-IV criteria, individual samples may still be phenotypically diverse, and thus potentially genetically diverse. Previous GWAS of AD have tried to address the latter issue by conducting more limited analyses with males only and subjects with early onset AD (63,64). Here we present a GWAS of AD, AD symptomatology, and a composite externalizing phenotype using an ethnically homogeneous, severely affected sample. In order to increase power, we elected to genotype all siblings from our original linkage sample that met criteria for AD instead of selecting one independent case from each family. Including affected siblings from multiply afflicted families may reduce locus heterogeneity.

Methods

Cases & Controls

Participants in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) were recruited in Ireland and Northern Ireland between 1998 and 2002. Further details of the study design, sample ascertainment, and clinical characteristics of this sample are described elsewhere (242). In brief, ascertainment of probands was mainly conducted in community alcoholism treatment facilities and public and private hospitals. Probands were eligible for study inclusion if they met the current

DSM-IV criteria for AD and if all four grandparents had been born in Ireland, Northern Ireland, Scotland, Wales, or England. After a prospective family was identified through probands, parents and potentially affected siblings whom the probands provided permission to contact were recruited. Probands, siblings, and parents were interviewed by clinically-trained research interviewers, most of whom had extensive clinical experience with alcoholism. The assessment included demographic characteristics, lifetime history of AD and other comorbid conditions, alcohol-related traits, personality features, and clinical records. All participants provided informed consent. We used a subset of the linkage sample for the current study, including 710 probands and affected siblings after filtering. The DSM-IV AD diagnosis was assessed in probands and siblings using a modified SSAGA (semi-structured assessment of the genetics of alcoholism) interview (version II,(142)). Items entered into our component score included ADsx, conduct disorder (CDsx), antisocial personality disorder (ASPDsx), drug dependence (DDsx) and novelty seeking scores (NS). Further details on these phenotypes can be found in Chapter 4 of this dissertation. We did not include ADHDsx in this principal component analysis, as we did in Chapter 4, because while there is evidence of genetic overlap between adolescent ADHD and adults AD (110), it is not as strong as for the other traits we used and AD.

1755 DNA samples from healthy, unpaid volunteers donating blood at the Irish Blood Transfusion Service from The Trinity College Dublin (TCD) were used as controls. These TCD Biobank controls were asked if they had any problems with alcohol and excluded if they endorsed this. These individuals were previously genotyped on Affymetrix v6.0 arrays (Santa Clara, CA) by the Wellcome Trust Case Control Consortium and the Broad Institute Center for Genotyping and Analysis as part of a GWAS of schizophrenia (258,259). Information about age and sex was available for these subjects.

Replication sample

Our replication sample consists of subjects from the European American (EA) portion of the COGA sample (64). Researchers ascertained cases (N=847 EA and N=345 AA) at alcohol treatment programs in 7 centers throughout the US (64), and lifetime AD diagnoses were made using DSM-IV criteria with information collected from the SSAGA (142). Controls (N=552 EA and N=140 AA) were recruited by the same centers from driver's license registries and dental clinics and were nondependent. Genotyping was conducted using Illumina HumanHap 1M BeadChips (Illumina, San Diego, CA). For more details about the sample and analyses, see Bierut et al. (65). We constructed an externalizing factor score in cases only using phenotypes that were mostly overlapping with the IASPSAD; however, the illicit drug variable (range=0-45) differed in that COGA used a sum of DSM-III dependence symptoms for the 5 drug classes available. Additionally, we chose not to include novelty seeking in our factor analysis because there was a high degree of missingness in the COGA data. Thus, 4 phenotypes from the COGA data were entered into a factor analysis, including ADsx, CDsx, ASPDsx, and DDsx.

Genotyping & Pre-calling QC

Genomic DNA from whole blood was required to pass rigorous quality standards prior to plating. Genotyping plates included probands and affected siblings, 11 sets of parents to check for Mendelian errors, duplicates, as well as a subset of TCD Biobank controls (N=102) to check for concordance with the other sites. Samples were individually genotyped at Vanderbilt's core facility on Affymetrix v6.0 SNP arrays. After obtaining allele signal intensities in the form of CEL files from Vanderbilt, we generated quality control (QC) metrics using both the apt-geno-qc

program within Affymetrix Power Tools (260) and Birdseed v2 (261). The apt-geno-qc program assesses one array at a time and only uses a small subset of SNPs, whereas Birdseed v2 provides QC metrics for arrays in aggregate. One of these metrics is Contrast QC (CQC), which assesses the ability of an experiment to categorize SNP signals into three genotype clusters. There are also separate CQC measures for both restriction enzymes, Nsp and Sty. All arrays with CQC < 0.04 and/or CQC Nsp/Sty < 0 were excluded. Additionally, apt-geno-qc and Birdseed v2 provide four different measures of sex, which had to be known and equivalent in order for the array to be included.

BEAGLECALL

After calling arrays using the Birdseed v2 algorithm and recognizing a site effect, the calling algorithm BEAGLECALL (BC) was implemented at Golden Helix, as this program has been successful in correcting batch effects in other data sets (Matt Keller, pers. comm.). BC takes not only allele signal intensity information into consideration, but also information about the genotypes of neighboring markers (262). The algorithm uses genotypic probabilities obtained from another program as a starting point. Since we performed initial QC in Birdseed v2, we used probabilities generated from this program as input. All CEL files from the three sites (Vanderbilt University, the Affymetrix Contract Facility, and the Broad Institute) were quantile normalized together using the Golden Helix SVS algorithm and called together. This includes case arrays from a separate GWAS for schizophrenia; the total set of arrays is referred to as the full array set. No sex chromosomes were entered into the analysis because BC is not able to call the Y-chromosome or X-chromosome in males. A total of 890,920 autosomal SNPs were entered into the BC algorithm. SNPs that did not conform to Hardy Weinberg equilibrium were removed at

each of the three iterations with increasingly stringent filters as follows: $p < 1 \times 10^{-10}$, $p < 1 \times 10^{7}$, $p < 1 \times 10^{-6}$. Genotypes were received from Golden Helix as probabilities and were converted into hard calls using a threshold of 98% probability of a specified genotype; otherwise, they were set to missing.

Post –BC SNP & Sample QC

We performed all QC checks in PLINK version 1.7 (251). Since the X-chromosome was not called using BC, we used Birdseed v2 calls to remove samples in which the phenotypic sex did not match the genotypic sex. The remaining QC steps were conducted using BC calls. We ran IBD estimation on a pruned subset of SNPs and generated pi-hats for each individual. Using the IBD estimates, we ran multi-dimensional scaling with HapMap3 samples, and samples that did not cluster tightly with each other toward the European HapMap3 samples were removed. Additionally, if either controls or case-control pairs were cryptically related (pi-hat > 0.05), one control from each pair was excluded. In order to maintain power, we did not remove two families with minimally elevated pair-wise pi-hats in the range of 0.07 to 0.11. Family members whose pi-hats did not match the pedigree file were re-coded according to the genotypic information and mean pi-hat outliers, representing DNA samples that were contaminated or poor quality, were excluded. Heterozygosity outliers, which may also represent poor quality samples and were highly overlapping with mean pi-hat outliers, were also removed. Graphs depicting the cutoffs for mean pi-hat and heterozygosity outliers are shown in Figure 5.1. Furthermore, for each duplicate pair, the sample with the higher call rate was carried forward. All samples had calls rates > 98%; therefore, none needed to be excluded based on this criteria. SNPs with $\geq$ 1 Mendel or duplicate error were removed. Additionally, we excluded SNPs with a low minor allele

frequency (<1%) and/or a high level of missingness (>2%) using the full array set. SNPs with significant differences ($p \leq 7.37 \times 10^{-8}$) in missingness between cases and controls within the alcohol subset were also removed.

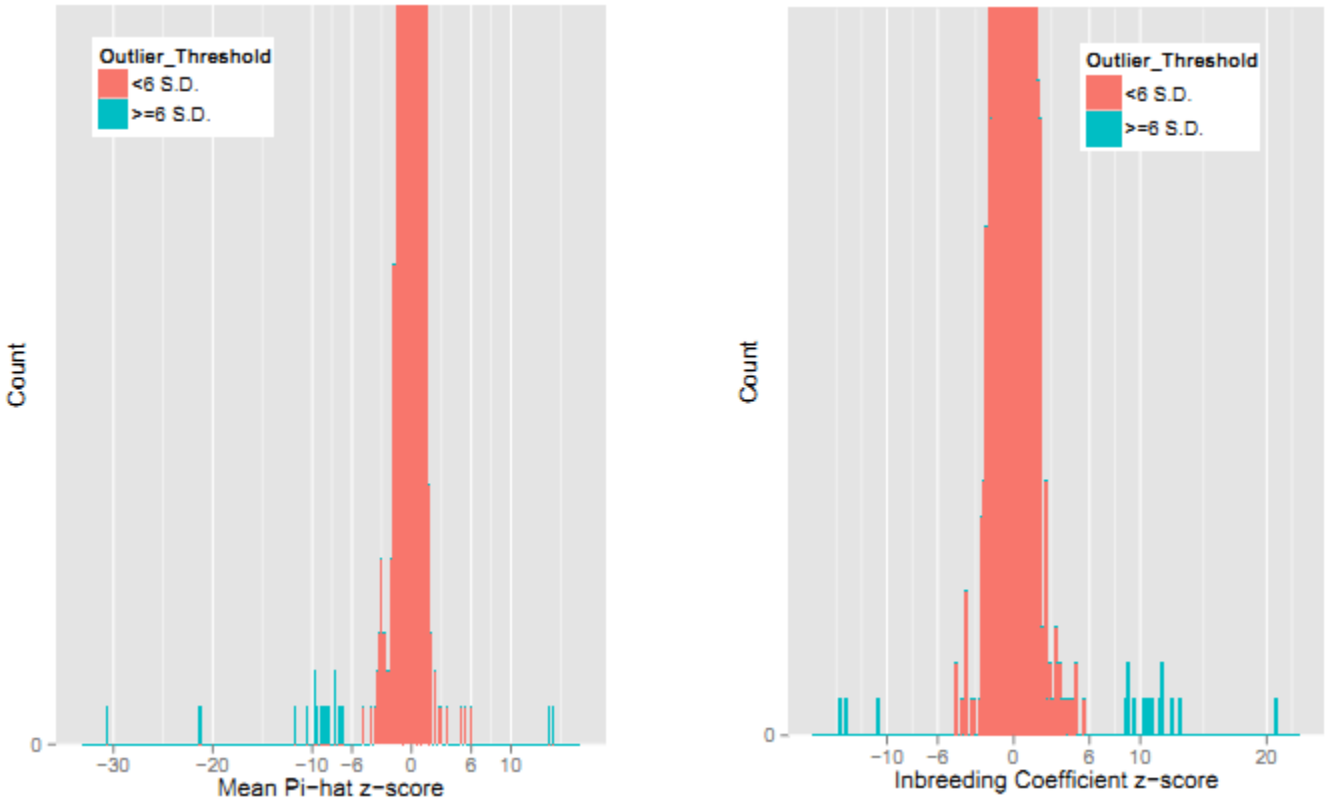

Figure 5.1. Histograms showing mean pi-hat and heterozygosity outliers.

Imputation

To increase coverage and comparability between the discovery and replication samples, we imputed the IASPSAD and COGA case control arrays (European American and African American samples) using IMPUTE version 2.2 (263) and the 1000 Genomes Project integrated variant reference panel from March 2012. This panel contains whole genome sequence from 1,092 individuals of African, American, Asian, or European Ancestry. Using samples from multiple ethnicities in the reference panel increases imputation accuracy (264). The panel

contains 36.5 million SNPs and 1.5 million structural variants and insertion/deletions, the majority of which are invariant in the Irish population. We filtered both the IASPSAD and COGA data sets post-imputation by excluding non-SNPs and SNPs with MAF < 1% and information criterion (an estimated correlation between imputed and true allele counts) ≤ 0.3, leaving 8.2 million and 8.5 million SNPs, respectively.

Statistical Methods

AD: Single Marker Analysis. We performed association tests on 710 cases and 1755 population controls using the Liming and Abecasis version of MQLS, which accounts for relatedness between subjects by using a kinship matrix calculated from pedigree data (265). We increased power and took advantage of additional phenotypic information from the full linkage sample by including relatives with phenotypic, but not genotypic, information in the analysis. MQLS takes these relatives into consideration by assigning individual weights based on the familial burden of disease. Controls were coded as having unknown phenotypes, as they did not undergo formal screening for AD. We used a sex weighted prevalence estimate of AD in our controls based on the National Epidemiological Study of Alcohol and Related Conditions report and data on alcohol consumption in non-remunerating blood donors from Holland (5,40). While no published studies have examined the prevalence of AD in this demographic, we were able to adjust the prevalence for the proportional difference in heavy drinking between blood donors and the general population, leaving us with an estimate of 8.9%. MQLS was run using allelic dosage data. IMPUTE2 outputs 3 posterior genotypic probabilities for each SNP, which can be converted to MACH style dosages using a command within the R program GenABEL (266). In

dosage format, each SNP is represented by a single number ranging from 0-2. MQLS does not allow for incorporation of covariates.

AD: Gene-based Analysis. We conducted gene-based analysis for both traits in the program Versatile Gene-based Association Study (VEGAS), which can be used with related individuals (267). All *p*-values were entered into the program and an empirical gene-wide *p*-value was generated based on SNPs lying within +/- 50 kb of a gene's 5' and 3' UTRs. The program converts all association *p*-values for SNPs in each gene region into chi-squared test statistics and sums them to obtain a gene-based test statistic. An empirical gene-based *p*-value is then derived via a simulation procedure in which the LD structure of the SNPs entered into the analysis are considered. The software uses the HapMap2 Caucasians as a reference population for determination of LD structure. Since we used 1000 Genomes for imputation, our haplotypic resolution was not as fine as it would have been if genotypes from this panel were used by VEGAS. Gene-based tests can provide different information than SNP-based association because a gene may be enriched for signal but no individual signals in this gene surpass a pre-defined threshold. One disadvantage of this method is that it devalues larger genes due to the large number of SNPs considered. We used two approaches to select markers and genes that we report here. First, we utilized a genome-wide significance threshold of $p \leq 3.06 \times 10^{-8}$, which controls the false positive rate (i.e. the rate at which truly null tests are called significant) at an alpha of $\leq$ 0.05 for 1.6 million independent markers in a European population (5,41). Second, we used a cutoff based on the false discovery rate (FDR), which is the rate at which tests called significant are truly null (43). To implement the latter method, for each *p*-value, we calculated a *q*-value (42,43), or an estimate of the proportion of false positives among all significant markers when the corresponding *p*-value is declared significant. We used a *q*-value threshold of 0.5 for

declaring significance (i.e. 50% of significant findings are false discoveries), since this threshold provides a good balance between controlling false positives and detecting true effects (268).

AD: Replication and Meta-analysis. We pursued three different replication/meta-analytic approaches to maximize the potential of our data. We used both the IASPSAD and COGA as discovery samples and attempted to replicate SNPs with $q \leq 0.5$ in each discovery sample. Additionally, to detect signals in which both the IASPSAD and COGA samples yield low *p*-values that may not have met the $q \leq 0.5$ threshold in either sample, we conducted a meta-analysis of all overlapping SNPs in each sample that passed post-imputation filtering within the meta-analytic program, METAL (269,270). Combining summary statistics within a meta-analytic framework is simpler than pooling data, especially when different analytic strategies, covariates, etc. are used within each sample and it is also just as efficient (271). This software provides the option of either combining test statistics and standard errors or *p*-values with consideration of sample size and direction of effect. We chose the latter method because the former requires that variables from different studies be measured in exactly the same way (272) and MQLS does not generate standard errors.

Quantitative Traits: Single Marker Analysis. We analyzed the quantitative traits ADsx and the externalizing component score in cases only using ProbABEL, which also utilizes a kinship matrix generated from genomic data to account for relatedness and allows for dosage data as input (266). ProbABEL uses a two-step mixed model-based score test in which the first step accounts for relatedness and the second step incorporates the effects of SNPs and covariates, which were age and sex.

Principal Components Analysis. Because the IASPSAD and COGA samples used different measures for some phenotypes (e.g. drug dependence symptoms) that we wished to

include in our factor analysis, we wanted to ensure that the factor scores based on these items captured the same underlying construct. To test whether this was so, we compared for each sample factor scores derived from non-overlapping but complete phenotypic information (i.e. those we wished to use for analyses) with factor scores derived from a reduced set of items present in both samples. Ideally, we wished to include a measure of symptoms for AD, ASPD, CD, and DD as well as novelty seeking scores in our factor analyses. Of these, AD, ASPD, and CD were assessed identically in both samples. The quantitative drug phenotype in COGA consisted of DSM-III-R symptoms for 5 drug classes, whereas the quantitative drug phenotype in the Irish sample consisted of the number of positive DSM-IV drug dependence symptoms across 7 drug classes. In the Irish sample, individuals had to endorse using drugs 11 times in the same month to qualify for the drug dependence questions, whereas in the COGA sample, subjects only needed to endorse use 11 times in a lifetime. Finally, COGA did not have enough individuals with NS scores to make inclusion of this trait in the analysis meaningful. Therefore, we entered phenotypes from each sample into the most complete factor analysis possible (i.e. making use of all externalizing phenotypes available, regardless of whether they were identical or not) as well as a more limited factor analysis based on the overlapping phenotypes assessed identically in both samples. The overlapping phenotypes included DSM-IV ADsx, DSM-III-R ASPDsx, and DSM-III-R CDsx as well as a drug phenotype consisting of the number of illicit substances tried. In all factor analyses, one factor could be retained by both eigenvalue and scree plot examination.

Within each sample, we calculated the complete component scores and the more limited overlapping component scores, and then generated Pearson correlations between these values. The procedure used to generate externalizing component scores was described in the previous

Chapter, but briefly, all phenotypes were entered into a principal component analysis and loadings were used to calculate individual component scores. The correlations between the more complete and the more limited component scores were 0.96 and 0.95 in the Irish and COGA samples, respectively. Thus, we concluded that the more detailed, though not completely overlapping factors scores, could be used for this analysis. In the Irish sample, the complete factor had an eigenvalue of 2.23 and explained 44.6% of the variance. The component loadings were as follows: ADsx, 0.37; ASPDsx, 0.85; CDsx, 0.87, DDsx, 0.61, and novelty seeking scores, 0.48. In the COGA sample, the complete factor had an eigenvalue of 1.94 and explained 48.7% of the variance. The component loadings for each phenotype were as follows: ADsx, 0.41; ASPDsx, 0.84; CDsx, 0.84; and DDsx, 0.70.

Results

AD: Single-Marker Analysis, Replication, and Meta-analysis

Quantile-quantile (QQ) and Manhattan plots for AD are shown in Figure 5.2. In the QQ plot, the distribution of *p*-values closely follows the null distribution except at the tail, which is expected if there is no site difference or population stratification and there are real effects of modest size. The lambda is 1.05. The Manhattan plot shows 13 SNPs met criteria for genome-wide significance ($p \leq 3.06 \times 10^{-8}$) with 12 falling in *COL6A3* in chromosome 2 and 1 falling in an intergenic region of chromosome 3. While genome-wide significance is an important consideration, as previously mentioned, it may be better to approach GWAS results using an FDR method; therefore, our results are filtered according to this criterion. The tables of results are arranged by chromosome and base pair to demonstrate, by assessment of MAF, that multiple

signals in the same region often represent non-independent signals. This is further supported by the LD relationships between signals in each region illustrated in the regional association plots for selected regions in Figures 5.5, 5.6, and 5.7. Table 5.1 shows the 180 SNPs with $q \leq 0.1$, while Table A.3 in the Appendix shows the 725 SNPs with q-value ≤ 0.5. Of these SNPs, 657 were present in the COGA sample and 31 showed evidence of replication at a nominal p-value; however, none showed evidence of replication by a $q \leq 0.5$ criterion.

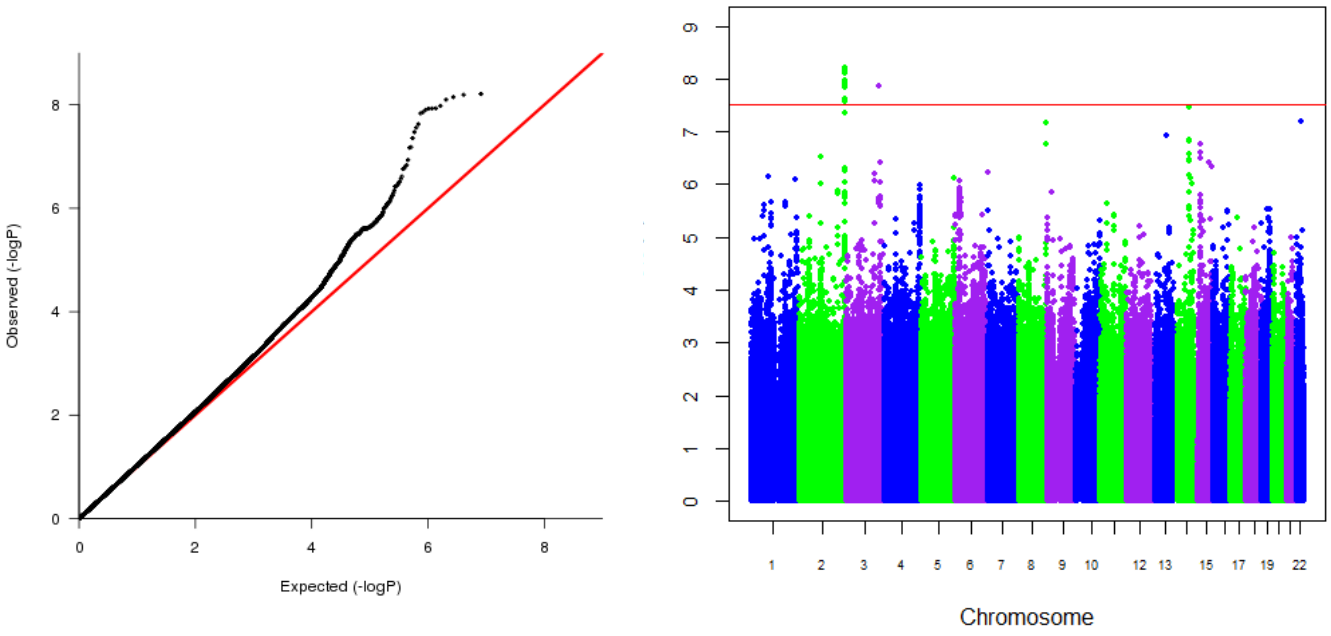

Figure 5.2. QQ and Manhattan plots of *p*-values for DSM-IV AD diagnosis. Analysis in 710 cases and 1755 controls. QQ plot shows observed vs. expected –log *p*-values. λ=1.05. Red line in QQ plot represents distribution under the null hypothesis of no association. Red line in Manhattan plot indicates genome-wide significance threshold at $p=3.06 \times 10^{-8}$.

Table 5.1. Association results for IASPSAD AD GWAS and COGA replication with *q*-values ≤ 0.1

SNP	CHR	BP	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE	LD ANNOTATION
rs72833664	2	114506030	3.07E-07	0.1	NA	NA	*SLC35F5*	-
rs56310758	**2**	**238240863**	**1.25E-08**	**0.1**	**0.04**	**0.6**	***COL6A3***	-
rs75561681	**2**	**238241881**	**2.77E-08**	**0.1**	**0.01**	**0.6**	***COL6A3***	-
rs2270671	**2**	**238243285**	**1.18E-08**	**0.1**	**0.06**	**0.6**	***COL6A3***	-
rs10929226	**2**	**238244559**	**1.17E-08**	**0.1**	**0.06**	**0.6**	***COL6A3***	-
rs112523013	**2**	**238247257**	**1.05E-08**	**0.1**	**0.07**	**0.6**	***COL6A3***	-
rs2646258	**2**	**238253149**	**6.39E-09**	**0.1**	**0.05**	**0.6**	***COL6A3***	-
rs2256485	**2**	**238253930**	**6.18E-09**	**0.1**	**0.03**	**0.6**	***COL6A3***	-
rs2646265	**2**	**238257013**	**1.46E-08**	**0.1**	**0.07**	**0.6**	***COL6A3***	-
rs2646264	**2**	**238257213**	**1.19E-08**	**0.1**	**0.10**	**0.6**	***COL6A3***	-
rs2646261	**2**	**238259387**	**2.36E-08**	**0.1**	**0.08**	**0.6**	***COL6A3***	-
rs11901326	2	238261509	5.00E-07	0.1	0.33	0.8	*COL6A3*	-
rs2646257	2	238261850	5.58E-07	0.1	0.33	0.8	*COL6A3*	-
rs2645764	**2**	**238262254**	**7.09E-09**	**0.1**	**0.18**	**0.6**	***COL6A3***	**-**
rs2645763	**2**	**238263299**	**7.97E-09**	**0.1**	**0.17**	**0.6**	***COL6A3***	**-**
rs2645777	2	238266146	4.41E-08	0.1	0.13	0.6	*COL6A3*	-
rs73149934	3	147209715	6.23E-07	0.1	0.53	0.9	intergenic	*ZIC4/ZIC1*
rs150268941	**3**	**172575312**	**1.39E-08**	**0.1**	**0.20**	**0.6**	**intergenic**	**-**
rs142645748	3	175451990	3.80E-07	0.1	NA	NA	*NAALADL2*	-
rs45619636	6	166823764	5.75E-07	0.1	0.66	1.0	*RPS6KA2*	-
rs79048468	8	137895174	1.74E-07	0.1	0.53	0.9	intergenic	-
rs117687198	8	137907672	6.78E-08	0.1	0.40	0.8	intergenic	-
rs117695261	13	74325506	1.16E-07	0.1	0.82	1.0	*KLF12*	-
rs117727648	14	77100157	3.36E-08	0.1	0.89	1.0	intergenic	-
rs192688395	14	77211718	2.70E-07	0.1	1.00	1.0	intergenic	-
rs113653607	14	77641605	3.61E-07	0.1	0.39	0.8	intergenic	-
rs142687658	14	77750946	1.47E-07	0.1	NA	NA	*POMT2*	-
rs150017190	14	77806492	3.47E-07	0.1	NA	NA	intergenic	-
rs56198483	14	77828430	1.53E-07	0.1	0.08	0.6	*TMED8*	*POMT2, GSTZ1*
rs4780153	15	33972420	1.69E-07	0.1	0.95	1.0	*RYR3*	-
rs1353348	15	33984848	3.19E-07	0.1	0.76	1.0	*RYR3*	-
rs939432	15	33986294	2.42E-07	0.1	0.81	1.0	*RYR3*	-
rs72742523	15	79713106	3.77E-07	0.1	0.94	1.0	intergenic	-
rs7183304	15	92420700	4.52E-07	0.1	0.79	1.0	*SLCO3A1*	-
rs185631468	22	42710609	6.56E-08	0.1	NA	NA	intergenic	-

Notes: Genome-wide significant signals in bold.

We show QQ plots for the *p*-values in the COGA discovery sample as well as the joint analysis meta-analysis in Figure 5.3. The lambdas for these analyses were 1.03 and 1.05, respectively. The COGA discovery component did not yield any *q*-values at the 0.5 level, so a table is not

presented. 196 SNPs with *q*-values $\leq$ 0.5 in the joint meta-analysis are presented in Table A.4. SNPs in all tables were annotated using BioQ, which utilizes information from dbSNP build 137 (273), while Table 5.1 has additional annotation information derived from consideration of LD in the program LocusZoom using hg19 European 1000 Genomes data (274). All neighboring genes that fell within an LD block defined by $r^2 \geq 0.2$ of the index SNP were also used to annotate the SNP.

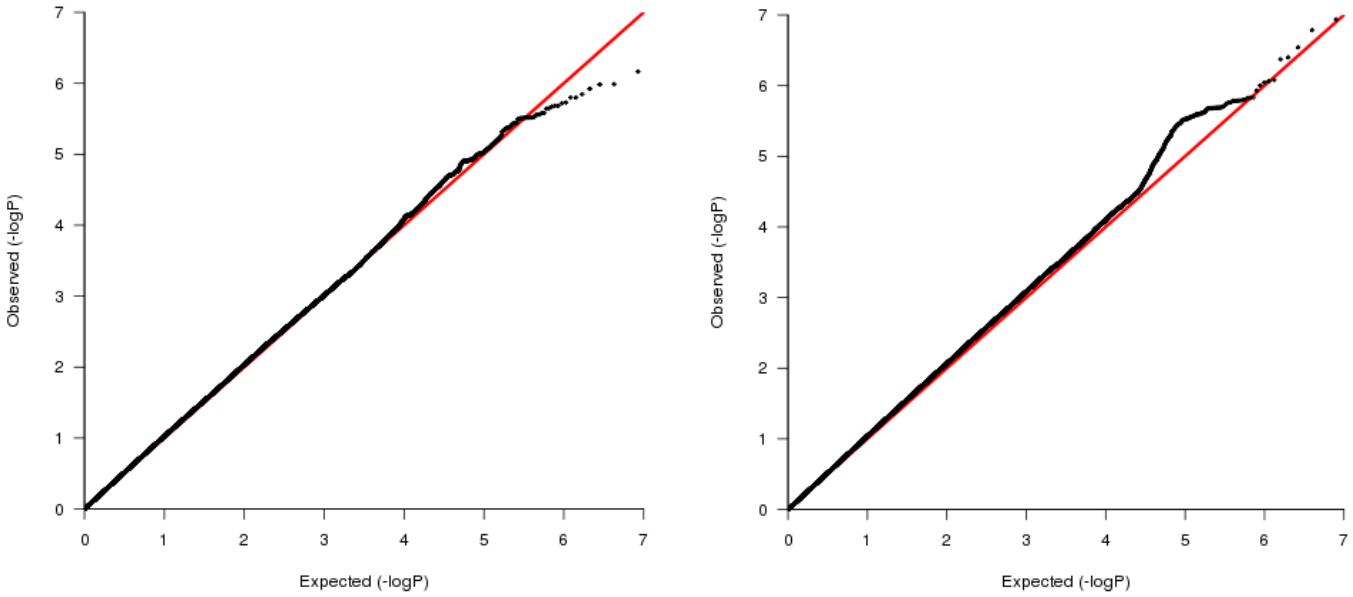

Figure 5.3. QQ plots of COGA and the joint meta-analysis *p*-values. Red lines represent distribution under the null hypothesis of no association. λ=1.03 and 1.05.

AD: Gene-based analysis

Genes with $q \leq 0.5$ in the VEGAS gene-based analysis are presented in the main text in Table 5.2. Some of the same genes present in Table A.3 also appear on the gene-based list, including *COL6A3*. One of the more promising findings from the SNP-based association

analysis, *RYR3*, is not on this list due to the fact that it is a very large gene (555 kb), and the signal is fairly focal. Therefore, VEGAS considered many non-significant SNPs in calculation of the gene-based statistic for this gene. Our list contains several clusters of genes in the same region that were likely generated from the same set of SNPs, supported by the fact that the top SNP is the same for many of these genes. This is more prone to happen with small genes. Additionally, there are several additional genes that were not present in the single marker analysis of AD in the IASPSAD data.

Table 5.2. VEGAS gene-based results for IASPSAD AD with *q*-values ≤ 0.5

CHR	GENE	nSNPs	START	STOP	GENE *p*	GENE *q*	BEST SNP
1	*MED8*	58	43622174	43628070	8.48E-04	0.5	rs2842182
1	*C1orf84*	64	43628142	43645152	7.09E-04	0.4	rs2782651
1	*KIAA0467*	73	43661383	43690891	6.37E-04	0.4	rs2105029
1	*HYI*	55	43689415	43692505	6.11E-04	0.4	rs2105029
1	*HSD3B1*	118	119851348	119859204	6.95E-04	0.4	rs1047303
1	*LEFTY2*	52	224190925	224195543	7.25E-04	0.4	rs2816330
2	*COL6A3*	230	237897393	237987589	7.00E-06	0.3	rs2256485
4	*UGT2B15-1*	8	69194909	69218969	2.89E-04	0.3	rs9999801
5	*LHFPL2*	253	77816793	77980404	1.44E-04	0.3	rs9293756
5	*UBXD8*	85	175807961	175869681	7.24E-04	0.4	rs2963296
6	*PRL*	151	22395458	22405709	1.29E-04	0.3	rs707870
6	*REV3L*	228	111726926	111911107	3.68E-04	0.3	rs455726
6	*TRAF3IP2*	185	111986835	112034014	3.09E-04	0.3	rs10872068
7	*MICALL2*	50	1440520	1465635	1.01E-04	0.3	rs10267348
18	*RAB27B*	114	50646837	50708209	2.70E-04	0.3	rs12953492
19	*HKR1*	67	42517419	42547197	3.03E-04	0.3	rs7258912
19	*ZNF527*	74	42553898	42575806	3.16E-04	0.3	rs7258912
19	*ZNF569*	87	42593899	42650179	2.40E-04	0.3	rs2161520
19	*ZNF570*	58	42651821	42668082	2.47E-04	0.3	rs2161520
19	*ZNF793*	84	42689680	42726079	1.17E-04	0.3	rs17245425
19	*ZNF540*	101	42734147	42796836	2.12E-04	0.3	rs2927743
19	*ZNF571*	82	42746994	42777513	3.02E-04	0.3	rs2927743
19	*ZFP30*	68	42815228	42838153	4.09E-04	0.4	rs2927743
19	*ZNF781*	68	42850489	42875056	5.09E-04	0.4	rs3095726
19	*ZNF607*	59	42879115	42902531	7.71E-04	0.4	rs3095726
19	*ZNF573*	81	42921040	42962040	6.06E-04	0.4	rs3095726
19	*CCDC61*	86	51190178	51213714	1.77E-04	0.3	rs4803895
19	*PGLYRP1*	85	51214280	51218163	7.10E-05	0.3	rs4803895
19	*IGFL4*	89	51234845	51236114	6.10E-05	0.3	rs4803895

ADsx and the Externalizing Component Score: Single Marker Analysis

QQ plots for ADsx and the externalizing component score are shown in Figures 5.4. The genomic inflation factors are 0.92 and 1.04, respectively. ADsx has limited variation in our sample (over 80% endorsed ≥ 6 symptoms), which may explain the deflated genomic inflation factor. While this trait has yielded interesting signals in the IASPSAD sample in previous work (23,275), it is clearly not ideal due the very limited variation in our severely affected clinical

sample. ADsx would be better assessed in a population sample or one that included controls, so that the full distribution of symptoms could be represented. There are several other quantitative traits that were measured in the IASPSAD, as discussed in Chapter 4, that may be more suited to gene discovery efforts that we plan to pursue. Nevertheless, one intergenic SNP did meet genome-wide significance for ADsx as well as one for the externalizing composite score. Table 5.3 shows the 5 SNPs with $q \leq 0.5$ for ADsx and Table A.4 shows the 95 SNPs with $q \leq 0.5$ for the externalizing component score.

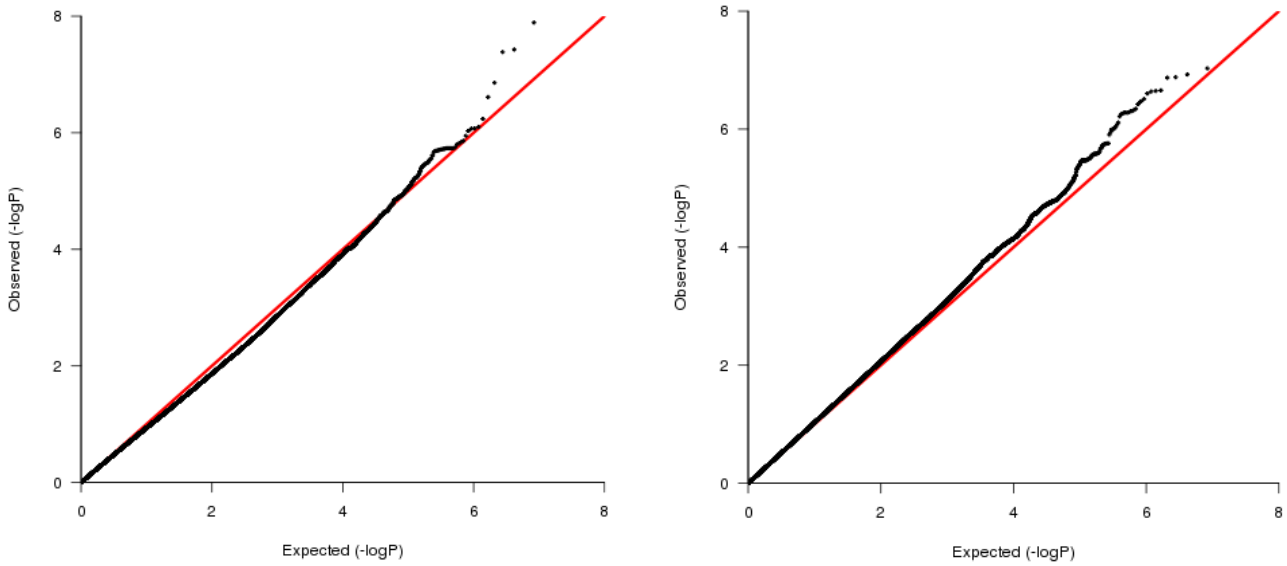

Figure 5.4. QQ plots of *p*-values for DSM-IV ADsx and the externalizing component score. Analyses in cases only. λ=0.92 & 1.04. Red lines represent distribution under the null hypothesis of no association.

Table 5.3: Association results for IASPSAD ADsx with *q*-values ≤ 0.5

SNP	CHR	BP	*p*	*q*	INFO	GENE	LD ANNOTATION
rs62563410	9	94554802	4.13E-08	0.4	0.60	*ROR2*	-
rs9557255	13	100213197	1.39E-07	0.4	0.94	*TM9SF2*	-
rs149373539	13	110351186	2.43E-07	0.5	0.49	intergenic	-
rs74473255	**18**	**52800791**	**1.29E-08**	**0.4**	**0.68**	**intergenic**	***TCF4***
rs73927900	19	15104035	3.73E-08	0.4	0.78	*SLC1A6*	*CCDC105*

Notes: Genome-wide significant SNPs in bold.

Discussion

Our AD GWAS of a clinically and ethnically homogeneous Irish sample yielded several potentially interesting and novel candidates. There was no evidence for replication in the COGA sample by a FDR criterion of $q \leq 0.5$; however, some genes showed nominal replication *p*-values ≤ 0.05. Furthermore, several genes present in Table A.3 in the Irish sample were also on the meta-analysis list (Table A.4), suggesting that the meta-analysis *p*-values were driven by the Irish data. However, there were also some new candidates, providing support for the utility of both a replication of top signals as well as a meta-analytic approach to the data. Many of these genes had low *p*-values (in the 10^{-6}-10^{-2} range) in both samples, but neither *p*-value was low enough to be on the top gene list in either sample.

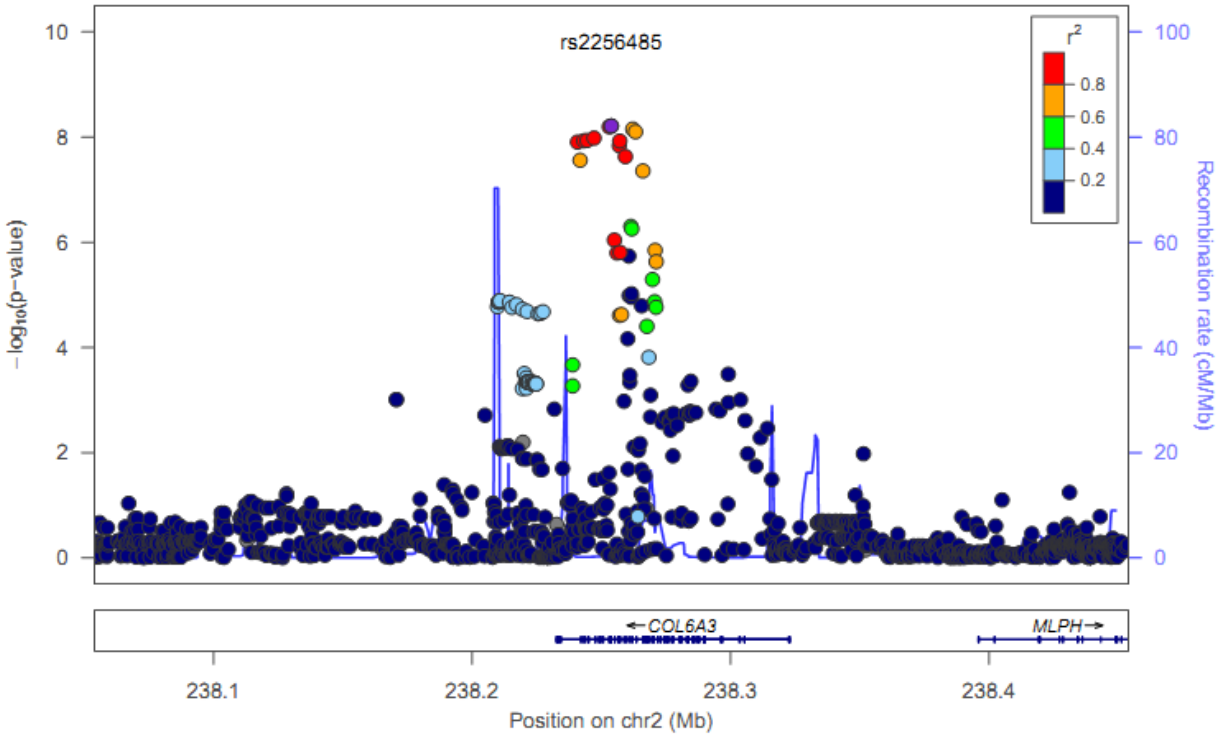

Figure 5.5. Regional association plot of top SNP in *COL6A3* and surrounding SNPs. Top SNP (rs2256485) is shown in purple and LD between that SNPs and other is indicated by colors shown in the legend.

There were 12 non-independent (as illustrated by LD relationships in the regional association plot in Figure 5.5) genome-wide significant signals in the novel candidate, *COL6A3*. This gene also harbors 4 SNPs with COGA replication *p*-values ≤ 0.05. *COL6A3* is one of six genes that encode α chains composing collagen VI, which is expressed in most connective tissues, including the extracellular matrices of the brain and liver. The most well-studied phenotypes associated with mutations in *COL6A3* are a range of congenital muscular dystrophies (CMD), including the milder Bethlem myopathy and the more severe Ullrich CMD (276). While some CMD include CNS involvement, the type IV collagen disorders do not. However, research suggests that collagen molecules function in neural cell migration, differentiation, and neurite outgrowth (277). Additionally, *COL6A3* was located under a linkage peak for an alcohol withdrawal factor score in a study using this sample. While this gene has not been implicated in alcohol previously, SNPs contained within another collagen gene, *COL8A1*, were among the ten top that appeared in the case-control COGA GWAS and replicated in the family sample (64).

In a gene expression study assessing mechanisms by which alcohol consumption enhances the progression of liver disease in patients with hepatic C virus, *COL6A3* showed a significant change in gene expression in cirrhotic livers (278). Considering our findings in light of this work, one might preliminarily hypothesize that there are polymorphism(s) in *COL6A3* that increase the risk of cirrhosis when the liver is exposed to high quantities of alcohol. To examine this more closely, we plan to pursue a follow-up association study in which we compare alcohol dependent cases who endorsed liver damage to those who did not. *COL6A3* was also found to be upregulated in colorectal tumor tissue (279). There is strong evidence for a causal link between heavy alcohol consumption and the subsequent development of cancer, especially

cancers of the digestive tract and female breast (4). Thus, it is possible that *COL6A3* may enhance risk for colorectal cancer development in individuals with heavy alcohol consumption.

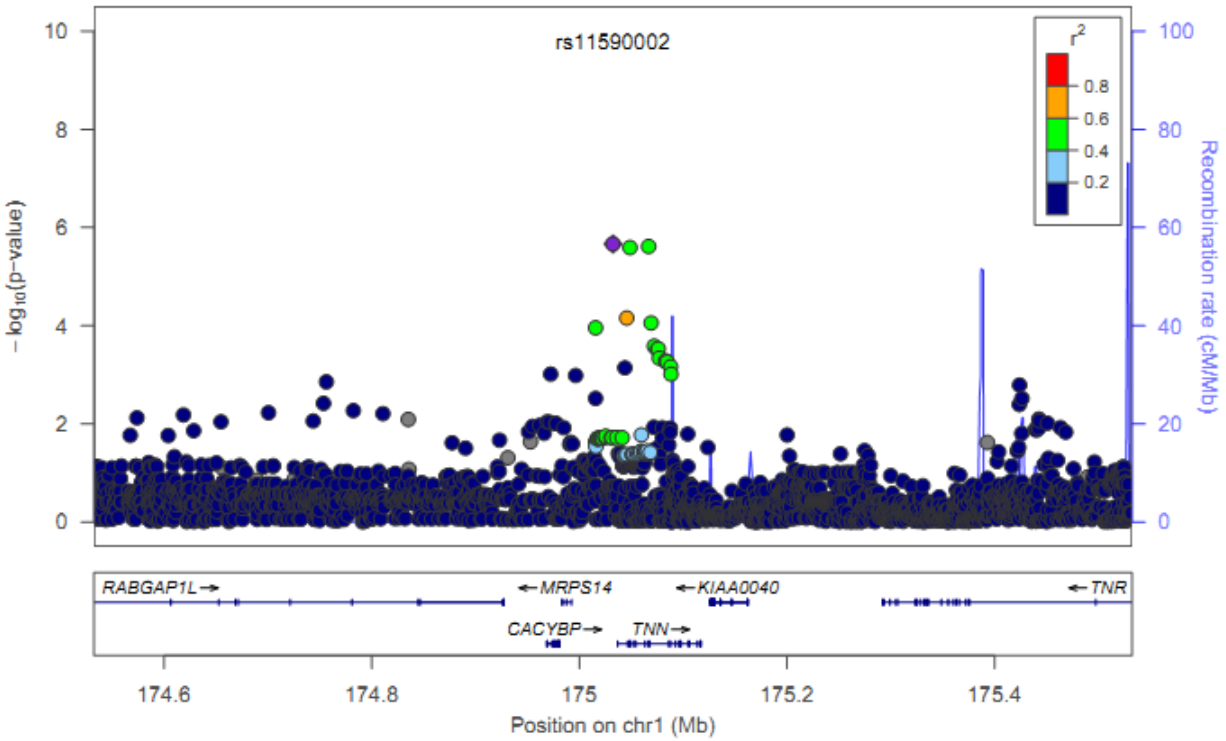

Figure 5.6. Regional association plot of top SNP near *TNN* and surrounding SNPs. Top SNP (rs11590002) is shown in purple and LD between that SNPs and other is indicated by colors shown in the legend.

TNN (also known as *TNW*), which encodes tenascin N/W, was present on both our top IASPSAD discovery and meta-analytic lists. Figure 5.6 shows the non-independent associated SNPs in the IASPSAD discovery sample. This gene is also related to cancer in that it is highly expressed in colorectal and female breast tumors (280). Additionally, it is induced by BMP2 (bone morphogenetic protein), which is a member of the BMP family and larger TGF-β family (281). There is some evidence that chronic alcohol exposure can affect BMP signaling in the liver of mice (282). Also, the BMP signaling pathway is related to another pathway involved in CLIC signaling, which are intracellular chloride channels that have been implicated in acute

ethanol behaviors in model organisms (283). Finally, in an analysis of combined case-control GWAS data from COGA and SAGE (the Study of Addiction: Genetics and Environment), Zuo et al. found that SNPs in the region of *TNN* and a neighboring gene, *KIAA0040*, were significant (284). No SNPs within *KIAA0040* were significant in our analysis.

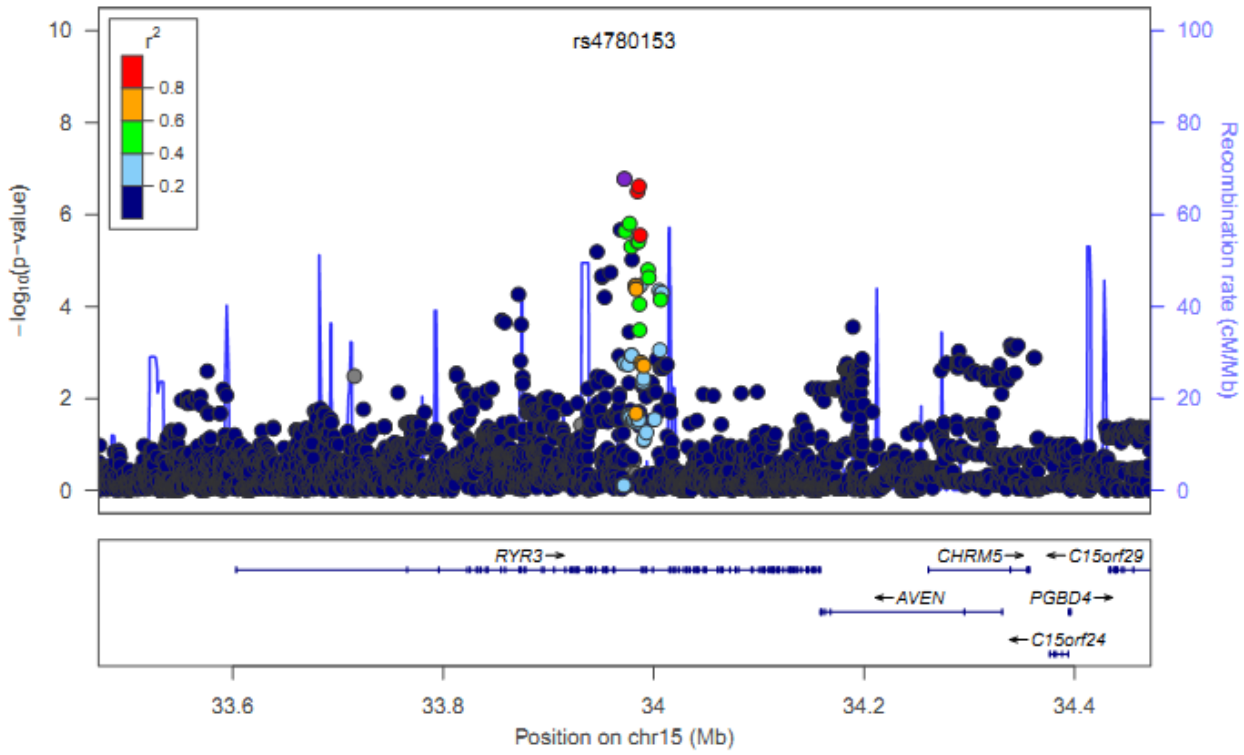

Figure 5.7. Regional association plot of top SNP in *RYR3* and surrounding SNPs. Top SNP (rs4780153) is shown in purple and LD between that SNPs and other is indicated by colors shown in the legend.

RYR3 is another interesting candidate due to its relationship to genes in the ethanol-related model organism literature. A regional association plot of associated SNPs (Figure 5.7) shows a focal signal toward the 5' end of the gene. *RYR3* encodes an intracellular calcium channel. Of the three ryanodine isoforms in humans, RyR3 is the most highly expressed in the brain (285). The ryanodine family of receptors has been shown to modulate BK channels (286), which were implicated in AD and alcohol-related traits in previous GWAS (64,68) as well as

model organism studies of ethanol behaviors (287). Furthermore, proteins in the ryanodine family have been found to interact with proteins in the CLIC family, intracellular chloride channels that have been implicated in acute ethanol behaviors in model organisms (283).

Several genes in Table 5.1 have been implicated in developmental delay and intellectual disability, including the linked genes *ZIC1* and *ZIC4 (*Zic family member 1 and 4), *NAALADL2* (N-acetylated alpha-linked acidic dipeptidase-like 2), *POMT2* (protein O-mannosyltransferase 2), and *SMARCA2* (WI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2) (present in A.3). *ZIC1* and *ZIC4* encode highly related zinc finger transcription factors; heterozygous deletions in this region are responsible for a common cerebellar birth defect called Dandy Walker Syndrome (288). *NAALADL2* is a candidate for Cornelia de Lange Syndrome, a disorder that involves neurological features, although no specific mutations were identified in a panel of patient DNA (289). Mutations in *POMT2* are known to cause a type of CMD, Walker-Warburg syndrome, which can involve developmental delay and mental retardation (290). Finally, *SMARCA2* is involved in regulation of gene transcription and heterozygous missense mutations have been shown to cause Nicolaides-Baraitser syndrome, which is characterized by facial and limb anomalies and intellectual disability (291). There is a phenotypic association with cognitive ability and AD/related traits as well as modest genetic correlations, as discussed in Chapter 2 of this dissertation (104). While the mutations and deletions previously described in these genes lead to gross neurological deficits, it is reasonable to hypothesize that more subtle variation may lead to brain changes that predispose to alcohol dependence.

Our gene-based analysis and joint meta-analysis for AD supported several interesting genes that were not implicated in our top results in either discovery sample. Of particular interest

in the gene-based analysis is *PRL*, which encodes the peptide hormone prolactin best known for its role in inducing lactation in mammals. However, this protein has a variety of other functions and was found to be elevated in alcohol dependent patients after alcohol intake (292) and while undergoing withdrawal (293). Interestingly, Schukit et al. (294) found that elevation in prolactin levels after alcohol consumption was related to a positive family history of AD, suggesting a possible genetic mechanism. Prolactin expression/secretion is influenced by both dopamingeric (295) and glutamatergic activity (296). Both neurotransmitters have been implicated in mediating the effects of alcohol. Another interesting series of genes from the VEGAS analysis was the set of zinc finger transcription factors, as *ZNF699* was previously associated in a candidate gene study of our sample (297). *ZNF699* is not in LD with any zinc finger transcription factors in Table 5.1, suggesting that multiple *ZNF* genes may be involved in AD. Genes to note from the joint analysis include *CSMD1*, which has been implicated in schizophrenia (298), and *CDH12*, which was a top hit in a GWAS of bipolar alcoholism (299).

While the single marker analyses for ADsx and the externalizing factor score yielded fewer SNPs with $q \leq 0.5$, some interesting genes were on these lists. The intergenic genome-wide significant hit for ADsx was in LD at the r^2=0.4 level with SNPs in *TCF4* (Transcription factor 4), which is involved in neurodevelopment and has been implicated in susceptibility to schizophrenia (300). This region is shown in Figure 5.8. A notable finding from the externalizing factor score results was in *GRID2* (Glutamate receptor, ionotropic, delta 2), which actually does not bind glutamate but is involved in synaptogenesis and synaptic plasticity in cerebellar cells (301).

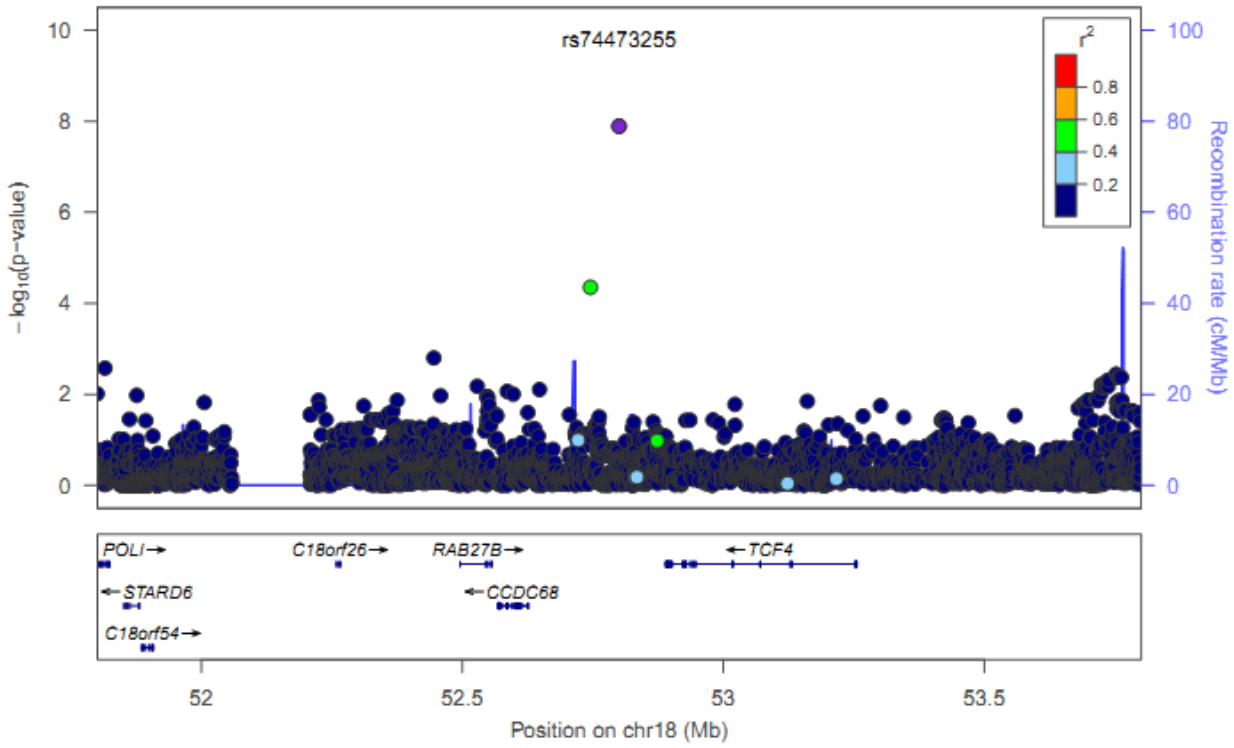

Figure 5.8. Regional association plot of top SNP near *TCF4* and surrounding SNPs. Top SNP (rs74473255) is shown in purple and LD between that SNPs and other is indicated by color.

Our results must be considered in the context of several limitations. First, the most extreme signals in any GWAS are also the most likely to represent errors/bias. We cannot fully discount the possibility that some or all of our signals are spurious; however, the extensive QC procedure and the use of BC calling algorithm would suggest that some of these signals are not due solely to bias or error. Second, for financial reasons, we used a sample of lightly screened controls, even though AD is a high prevalence condition. However, we accounted for this by coding the controls as unknown and using a prevalence estimated based on quantity of consumption in non-remunerating blood donors (40). Third, we analyzed the binary diagnosis of AD because it is a standard measure with good reliability that has been used in several other genome-wide association studies, allowing for direct comparison with replication samples. However, as discussed in the Introduction to this dissertation, we know from twin studies that it

is a genetically heterogeneous condition, which can reduce power to detect associated loci. For this reason, we also examined two quantitative traits, including ADsx and an externalizing component score.

CHAPTER 6: General Discussion

Part I Summary, Future Directions, and Limitations

The studies presented in this dissertation have attempted to expand our understanding of alcohol use disorders and related phenotypes with the use of two genetically informative study designs: twin (Part I) and association studies (Part II). In Chapter 2, we showed that education level is moderately genetically correlated with maximum drinks in 24 hours. Furthermore, we found that education significantly moderated the heritability of alcohol traits such that genetic influence was enhanced with increasing levels of education. These results may be useful in future work in two ways. First, the robust genetic correlation between education level and ARPs suggests that heritability estimates of these conditions are influenced by educational attainment. If twin modelers are interested in more "refined" phenotypes, they may consider using education level as a covariate in their twin analyses. For example, Agrawal et al. (302) found the heritability of regular cigarette smoking to be lower when they accounted for significant covariates in their twin model. Likewise, researchers conducting GWAS of ARPs could incorporate educational attainment into their analysis if they were interested in genes that influence a more refined alcohol phenotype. Also, the moderation findings could be used to inform forthcoming GWA studies in which education level is measured in the subjects. Some genes may not have main effects on risk of AD or ARPs and, therefore, will only be detectable at the extremes of the educational spectrum.

We presented two main findings in Chapter 3. First, we showed that, compared to wine drinkers, beer and spirits preference drinkers are at greater risk of higher levels of consumption, AUD diagnoses, and particular AUD symptoms, as well as ASPD and illicit substance

abuse/dependence. Together with findings from the extant literature, these results suggest that spirit preference drinkers may be more likely to belong to a chronic/severe subtype of alcoholics consisting of early onset AD and a high risk for antisocial behavior and polysubstance use. Our findings also indicate, in contrast, that beer preference drinkers are more liable to be part of a subtype with class with high levels of drinking and legal consequences, but less probability of comorbidity. Characterizing alcohol dependent individuals into meaningful groups may be helpful in terms of treatment efforts and gene discovery. Second, we found that total AC measures index genetic liability to preference to a modest degree, indicating that association analyses for these total measures are not likely particular useful in identifying genetic influences on intake of particular beverages. Identifying the genes may be helpful in that they could also be influential in characterizing individuals into specific typologies by altering disease characteristics such as age of onset, severity, course of illness, and symptom profile (185).

The findings from Part I of this dissertation must be interpreted with several limitations in mind. In Chapter 2, we were not able to use all of the same measures as in the Finnish study because RAPI scores were not assessed in the VATSPSUD; however, these scores were correlated at 0.55 with alcohol dependence symptoms in a subsample of the Finnish twins (73). Additionally, lifetime symptoms of ADsx and maximum consumption during the heaviest year were based on participants' retrospective recall and it is possible that our results may have been influenced by recall errors. In Chapter 3, we did not ask subjects directly about their alcoholic beverage preference, but rather assumed that their most frequently/heavily consumed beverage would reflect their preferences. Although alcoholic beverage choice may not always reflect preference, Straus and Bacon showed that the two traits are highly correlated (199). Also in Chapter 3, we did not have the data to assess additional alcoholic beverages beyond the three

main classes, including alcopops, cider, and caffeinated alcoholic beverages, which have been shown to have differential associations with particular traits as compared to wine, beer, and spirits (123,205). Finally, for both epidemiological studies, it is unclear whether the samples of Virginia- and Finnish-born Caucasians generalizes to other populations.

Part II Summary, Future Directions, and Limitations

In Chapter 4, we utilized a candidate gene design and found some evidence of involvement of particular DA genes in AD, ARPs, and externalizing behavior. The fact that we found a locus associated with general externalizing behavior in addition to individual phenotypes underscores the need for more association studies that consider the genetic overlap between phenotypes. In Chapter 5, we employed a genome-wide approach and identified several novel candidate genes for AD. 12 non-independent signals within *COL6A3* were genome-wide significant. Other top findings by $q \leq 0.1$ fell within several categories, including those involved in neurodevelopmental disorders and cancer. Besides being false positives, there several other reasons why none of the other top markers showed evidence of replication. These include that our associated alleles are specific to an Irish or a severely affected population. Despite the lack of replication, several of our top signals are promising candidates based on prior model organism and molecular literature. Thus, follow-up studies should still be pursued. Results from Part II must be viewed within the context of several limitations. First, our strongest SNPs are also the most likely to represent errors/bias; however, we employed an extensive QC process to limit the effects of error. Second, for financial reasons, we used a sample of population controls in our GWAS, even though AD is a high prevalence condition. However, we accounted for this in the binary analysis by coding the controls as unknown and using a prevalence estimated based on

quantity of consumption in non-remunerating blood donors (40). Third, AD is not an ideal trait for gene discovery. For this reason, we also examined additional quantitative traits in both Chapters 4 and 5.

Post-GWAS analysis. There are multiple ways to follow-up on these findings to gain more insight into their validity and functional relevance. One step will be to attempt replication in additional samples; however, even a real result may not replicate due to variety of factors, including lack of power in the replication sample, differences in the phenotypic presentations of the samples that have different underlying genetic causes, or ethnic differences between the samples. Even loci are replicated in other samples, it would not be clear whether the top SNPs are functionally relevant, or more likely, in LD with a functionally relevant SNP. At this point, bioinformatics tools or sequencing may be used to identify potentially causal SNPs to carry forward to molecular analyses. Because we imputed to the most recent 1000 Genomes reference panel, we will have a good handle on the common variants that are in high LD with the associated SNP. These SNPs may be evaluated for functional relevance by *in silico* examination of genomic location, influence on gene expression or protein function, and previous association with other phenotypes (303). Various molecular follow-up strategies could then be applied to prioritized SNPs based on predicted function. The most straightforward possibility is an exonic SNP that is predicted to affect protein function, although these are rarely associated with complex diseases (304). The only known SNPs of this kind for AD occur in the alcohol metabolizing genes and are generally protective. More common possibilities include SNPs located in known regulatory regions, such as enhancers, silencers, microRNA (short RNAs that cause gene silencing) binding sites, or non-coding regulatory RNA genes. In these cases, the effect of the SNP on regulation must be confirmed and then putative target genes must be

identified. One of the most common and challenging possibilities to interpret is a SNP in an intergenic regions or gene desert (304). One option in this case is to conduct targeted sequencing to detect SNPs within neighboring genes that may be driving the association. Both empirical and computational data provide evidence that a good proportion of trait-associated loci will contain variants, called expression quantitative trait loci (eQTLs), that influence the abundance of specific transcripts (305). We have access to postmortem brain samples from the New South Wales Tissue Resource Center that can be used to evaluate expression levels for genes potentially regulated by a top SNP.

Model Organisms. To evaluate their functional relevance specifically in ethanol related behaviors, prioritized SNPs/genes may be tested in model organisms. Because both invertebrates and vertebrates encounter fermenting fruits in their natural environments, they have developed physiological processes for metabolizing ethanol. Many organisms show similar reactions to ethanol as humans, including intoxication, tolerance, and withdrawal (306) and utilize similar molecular and physiological pathways underlying ethanol-related behaviors (307). Not only are these organisms useful for following up on specific loci/genes, they are also helpful in evaluating the role of networks to which the genes belong in alcohol related phenotypes. Within the VCU ARC, we have a variety of model organisms to test our top findings, each with its own advantages. The majority of ethanol related behaviors assessed in these organisms are acute; that is, they measure reaction to alcohol on a temporary basis. This may be relevant to certain human phenotypes, such as initial sensitivity, but with the addition of the Rat core to our ARC, we will be able to assess more chronic phenotypes that are generally more in line with our human phenotypes.

Rare Variation. Some have argued that despite the great deal of money and effort devoted to GWAS, much of the heritability for most complex traits remains to be explained and this missing heritability is likely due to rare variation (308,309). While it is true that the collective genetic variance explained by GWAS findings is still relatively small compared to the heritabilities estimated by twin studies, several lines of evidence suggest that the remaining heritability can mostly be explained by common variants of small effect and not rare variants of larger effect (310). Whole-genome analysis involving GWAS data of several psychiatric disorders, including schizophrenia and bipolar disorder, suggest that a good proportion of heritability estimated from twin studies can be explained by common variation (311). The variants that have surpassed genome wide significance for psychiatric disease thus far represent the largest effect sizes to be discovered. The others are obscured by the exceptionally high thresholds of significance. Other arguments for the so-called infinitesimal model include that the allele frequency distributions in GWAS are consistent with a limited number of rare variants, sibling recurrence rates are greater than would be expected for the postulated effect sizes of rare variants, and common variants are consistent across populations.

Nevertheless, most in the field would agree that studies of rare variation should be pursued to explain some proportion of the missing heritability. There has been good success with finding both common and rare CNVs associated with certain psychiatric disorders, including autism, bipolar disorder, and schizophrenia (312). CNVs in addiction have started receiving attention more recently. To our knowledge, the first studies of CNVs in relation to AD and ARPs were published in late 2011 and 2012 and include three on AD and alcoholism (313-315), one on amount of consumption (316), and one on brain volume in AUDs (317). These studies have implicated regions that are both duplicated and deleted and contain genes involved in

neurological dysfunction (313), developmental processes, cell communication, and sensory perception (315) , and enzymatic activity (316). Sequencing studies have now been published for several psychiatric traits and disorders, including autism, schizophrenia, and bipolar disorder and lend support to the idea that the same genes may harbor both common and rare variation that contributes to disease (318). We plan to follow-up on top genes from our GWAS in the same sample by assessing rare variation using a high pass targeted sequencing approach. Furthermore, we plan to follow-up top loci using low pass sequencing in a community sample of college students, which will harbor greater variation in quantitative traits than our severely affected clinical sample.

Lessons on Study Design

It is now commonplace in psychiatric genetics to combine samples that were collected and/or genotyped at different sites both in individual studies, as was the case in our GWAS, as well as through consortia. Even if genotyping occurs at the same site, batch effects can arise when investigators are not careful about randomizing cases and controls on different plates. While non-differential error between cases and controls results in reduced statistical power, the more serious type is differential error, which can lead to false positives. Differential error or missingness may be caused by a variety of differences in laboratory conditions, including variable technical skill and hybridization temperature of DNA to chips as well as use of different lots of reagents/chips. Ideally, researchers would collect and genotype cases and controls at the same site in randomly distributed batches during the same time period. However, for practical and financial reasons, this is often impossible. Therefore, investigators with suspected site or batch effects must dedicate sufficient effort to dealing with these artifacts to avoid clouding the

literature with false positives, such as in the example of the manuscript on a GWAS of longevity that was withdrawn from *Science* (319). Even worse, there was a case in which a diagnostic assay for ovarian cancer was put on the market but it was later revealed that the biological variable of interest was highly correlated with processing day (320). There are several ways to confirm the existence of site or batch effects. We conducted multidimensional scaling of identity-by-descent estimates by site and found that each generated its own distinct cluster. We also ran an association analysis with site as the phenotype in controls only (from Affymetrix and Broad) and found severe elevation in the genomic inflation factor. If site differences are confirmed and are not completely confounded with the outcome of interest, several analytic techniques may be helpful. One common technique is to use known technical variables as covariates, such as site and processing date. However, this strategy may not account fully for unmeasured non-biological variables, such as laboratory conditions, reagent quality, and personnel changes. Another strategy is called surrogate variable analysis (SVA), and was originally developed for gene expression microarrays, which are highly prone to be influenced by technical variables (321). In this method, unmeasured variables are estimated from the data and treated as standard covariates.

Since the processing laboratories were completely correlated with affection status in our situation (e.g. cases genotyped entirely at one site and controls at two others), we needed to pursue other approaches. Here we outline the process we implemented as a guide to other researchers dealing with technical artifacts in GWAS. Although our initial process did not correct the issue, there was a significant reduction that may be sufficient for researchers dealing with a less severe site effect. A standard step prior to data analysis is normalization, which adjusts global properties of samples to compare them more appropriately. After experimenting

with different combinations of normalization and calling, we found that normalizing and calling all samples separately most reduced the site difference. We called the data using Birdseed v2 (261), which uses prior information from HapMap3 as well as signal intensity data to cluster individuals into three groups according to their genotype. We then proceeded through a series of quality control steps in an effort to remove samples and SNPs that were responsible for the greatest amount of differential error between sites. We utilized all standard GWAS QC steps, such as removing arrays that were ancestry and heterozygosity outliers, but we also attempted several novel and under-used approaches. After an initial round of calling and QC, we recalled the samples because it has been shown that even a few poor quality samples can "profoundly" affect the allele calls of good quality samples (322).

Figure 6.1 shows quantile-quantile (QQ) plots of *p*-values for an association analysis in controls only using site as the phenotype. The X-axis shows the expected distribution of *p*-values under that the null hypothesis of no association, while the Y-axis shows the distribution of *p*-values for the analysis. The red line represents the distribution under the null in which the expected equals the observed (x=y), while the black points illustrates *p*-values generated in the analyses. In GWA studies with our sample size and expected effect size, the black points should follow the line until near the end, indicating mostly negative results with a handful of potentially true positive results at the tail. The first panel shows an analysis after Birdseed calling and QC. The line of points deviates early and sharply from the expected line, suggesting systematic differences between controls, such as population stratification or site effects. We knew there was not population stratification in our sample from previous tests, so the inflation represents a severe site effect. The genomic inflation factor (λ) quantifies the amount of deviation from the null. A λ of 1 would represent no deviation, but the first plot is highly inflated at 1.29.

At this point, we were fortunate to come across BEAGLECALL in the literature (262). This program was able to eliminate the site difference because calls are not only based on allele intensities, which were highly variable by site, but also LD information from neighboring SNPs. In the remaining panels of Figure 6.1, λ progressively declines with each iteration of BEAGLECALL until it reaches 1.06 in the last panel. As a field, it is imperative that we properly address the consequences of imperfect study design as we move forward with more advanced technologies, such as next generation sequencing (NGS). NGS has been subject to the same site and batch effects that plague other high throughput technologies, as seen in analyses of data from the 1000 Genomes project (320,323). Nevertheless, tools are being developed to deal with these artifacts, but researchers must be educated in their use (323).

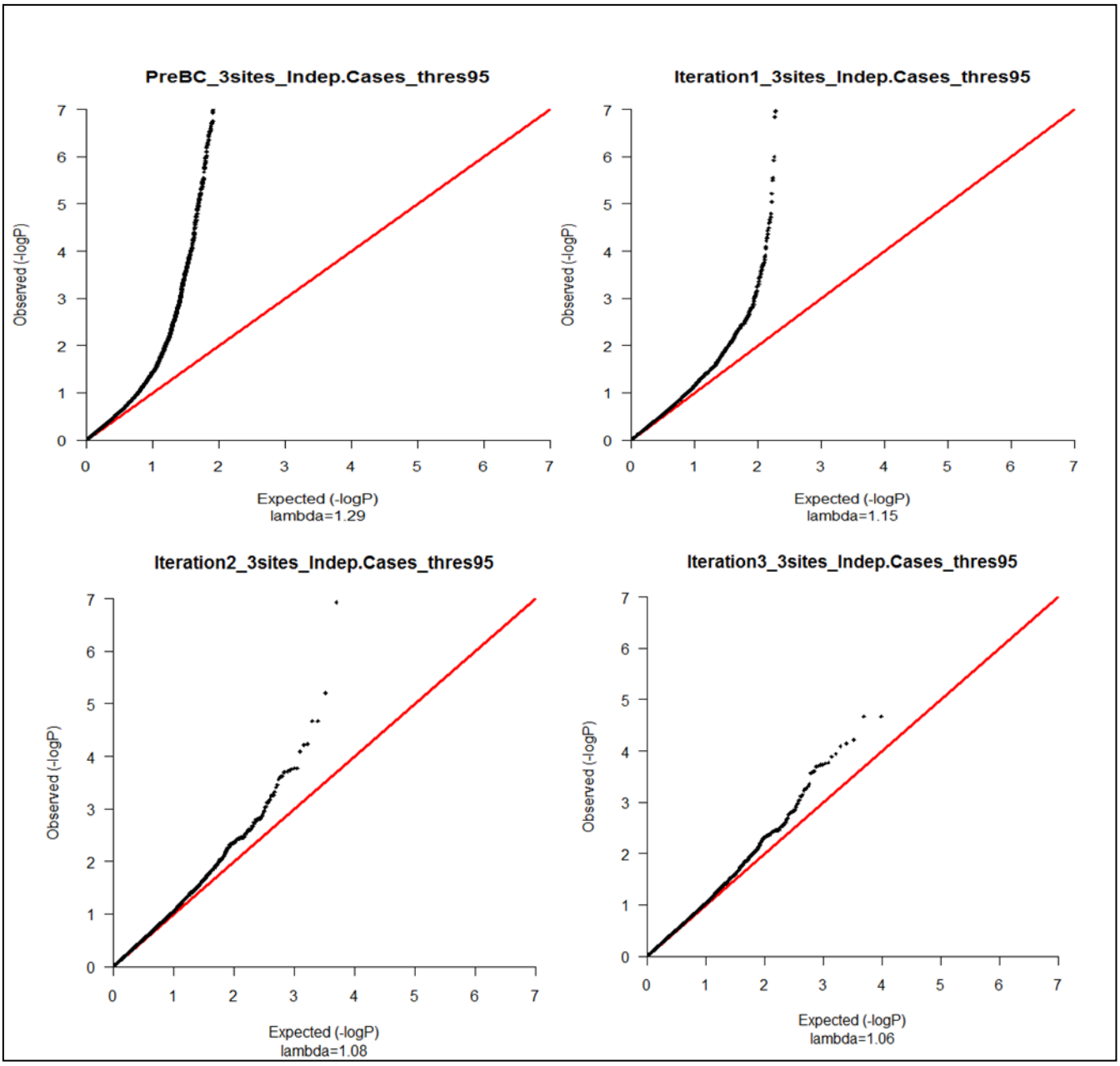

Figure 6.1. QQ plots of a control-only analysis using site as a phenotype. The first panel shows a QQ plot of the association analysis using data called with Birdseed v2, while the remaining panels show QQ plots of the various iterations with BeagleCall.

Literature Cited

(1) Miller WR, Wilbourne PL. Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. Addiction 2002 Mar;97(3):265-277.

(2) Caputo F, Trevisani F, Bernardi M. Alcohol misuse and traffic accidents. Lancet 2007 Feb 10;369(9560):463-464.

(3) Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. Lancet 2009;373:2223-2233.

(4) Rehm J. The risks associated with alcohol use and alcoholism. Alcohol Res Health 2011;34(2):135-143.

(5) Hasin DS, Stinson FS, Ogburn E, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch Gen Psychiatry 2007;64:830-842.

(6) American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.

(7) Prescott CA, Madden PA, Stallings MC. Challenges in genetic studies of the etiology of substance use and substance use disorders: introduction to the special issue. Behav Genet 2006;36:473-482.

(8) Dick DM, Prescott C, McGue M. The genetics of substance use and substance use disorders. Handbook of Behavior Genetics New York: Springer Science+Business Media, LLC; 2009. p. 435-455.

(9) Kendler KS, Myers J, Dick D, Prescott CA. The relationship between genetic influences on alcohol dependence and on patterns of alcohol consumption. Alcohol Clin Exp Res 2010 Jun;34(6):1058-1065.

(10) Dick DM, Meyers JL, Rose RJ, Kaprio J, Kendler KS. Measures of current alcohol consumption and problems: two independent twin studies suggest a complex genetic architecture. Alcohol Clin Exp Res 2011 Dec;35(12):2152-2161.

(11) Hansell NK, Agrawal A, Whitfield JB, Morley KI, Zhu G, Lind PA, et al. Long-term stability and heritability of telephone interview measures of alcohol consumption and dependence. Twin Res Hum Genet 2008 Jun;11(3):287-305.

(12) Kokkevi AE, Arapaki AA, Richardson C, Florescu S, Kuzman M, Stergar E. Further investigation of psychological and environmental correlates of substance use in adolescence in six European countries. Drug Alcohol Depend 2007 May 11;88(2-3):308-312.

(13) Kendler KS, Neale MC, Prescott CA, Kessler RC, Heath AC, Corey LA, et al. Childhood parental loss and alcoholism in women: a causal analysis using a twin-family design. Psychol Med 1996 Jan;26(1):79-95.

(14) Kuo PH, Gardner CO, Kendler KS, Prescott CA. The temporal relationship of the onsets of alcohol dependence and major depression: using a genetically informative study design. Psychol Med 2006 Aug;36(8):1153-1162.

(15) Kendler KS, Heath AC, Neale MC, Kessler RC, Eaves LJ. Alcoholism and major depression in women. A twin study of the causes of comorbidity. 1993;50(9):690-698.

(16) Kendler KS, Aggen SH, Prescott CA, Crabbe J, Neale MC. Evidence for multiple genetic factors underlying the DSM-IV criteria for alcohol dependence. Mol Psychiatry 2011 Nov 22.

(17) Dick DM. Gene-environment interaction in psychological traits and disorders. Annu Rev Clin Psychol 2011 Apr;7:383-409.

(18) Goldman D, Ducci F. Deconstruction of vulnerability to complex diseases: enhanced effect sizes and power of intermediate phenotypes. ScientificWorldJournal 2007 Nov 2;7:124-130.

(19) Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 2003 Apr;160(4):636-645.

(20) Agrawal A, Grant JD, Littlefield A, Waldron M, Pergadia ML, Lynskey MT, et al. Developing a quantitative measure of alcohol consumption for genomic studies on prospective cohorts. J Stud Alcohol Drugs 2009 Mar;70(2):157-168.

(21) Wood AC, Neale MC. Twin studies and their implications for molecular genetic studies: endophenotypes integrate quantitative and molecular genetics in ADHD research. J Am Acad Child Adolesc Psychiatry 2010 Sep;49(9):874-883.

(22) Hack LM, Kalsi G, Aliev F, Kuo PH, Prescott CA, Patterson DG, et al. Limited Associations of Dopamine System Genes With Alcohol Dependence and Related Traits in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD). Alcohol Clin Exp Res 2010 Nov 17.

(23) Kalsi G, Kuo PH, Aliev F, Alexander J, McMichael O, Patterson DG, et al. A systematic gene-based screen of chr4q22-q32 identifies association of a novel susceptibility gene, DKK2, with the quantitative trait of alcohol dependence symptom counts. Hum Mol Genet 2010 Jun 15;19(12):2497-2506.

(24) Foroud T, Wetherill LF, Liang T, Dick DM, Hesselbrock V, Kramer J, et al. Association of alcohol craving with alpha-synuclein (SNCA). Alcohol Clin Exp Res 2007 Apr;31(4):537-545.

(25) Wang JC, Hinrichs AL, Bertelsen S, Stock H, Budde JP, Dick DM, et al. Functional variants in TAS2R38 and TAS2R16 influence alcohol consumption in high-risk families of African-American origin. Alcoholism Clinical and Experimental Research 2007;31:209-215.

(26) Eisenberg N, Sadovsky A, Spinrad TL, Fabes RA, Losoya SH, Valiente C, et al. The relations of problem behavior status to children's negative emotionality, effortful control, and impulsivity: concurrent relations and prediction of change. Dev Psychol 2005 Jan;41(1):193-211.

(27) Hinshaw SP. On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. Psychol Bull 1987 May;101(3):443-463.

(28) Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. Arch Gen Psychiatry 2003;60(9):929-937.

(29) Kessler RC, Adler L, Barkley R, Biederman J, Conners CK, Demler O, et al. The prevalence and correlates of adult ADHD in the United States: results from the National Comorbidity Survey Replication. Am J Psychiatry 2006;163:716-723.

(30) Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, McGue M. Etiologic connections among substance dependence, antisocial behavior, and personality: modeling the externalizing spectrum. J Abnorm Psychol 2002;111(3):411-424.

(31) Dick DM, Aliev F, Wang JC, Grucza RA, Schuckit M, Kuperman S, et al. Using dimensional models of externalizing psychopathology to aid in gene identification. Arch Gen Psychiatry 2008;65(3):310-318.

(32) Kendler KS, Jaffee S, Romer D editors. The Dynamic Genome and Mental Health: The Role of Genes and Environments in Youth Development. 1st ed. USA: Oxford University Press; 2011.

(33) Kendler KS. Psychiatric genetics: a methodologic critique. Am J Psychiatry 2005;162(1):3-11.

(34) Dick DM, Kendler KS. The impact of gene–environment interaction on alcohol use disorders. Alcohol Research: Current Reviews ;34(3):318-324.

(35) Duncan LE, Keller MC. A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. Am J Psychiatry 2011;168(10):1041-1049.

(36) Dick DM, Latendresse SJ, Riley B. Incorporating genetics into your studies: a guide for social scientists. Front Psychiatry 2011;2:17.

(37) Sanghera DK, Blackett PR. Type 2 Diabetes Genetics: Beyond GWAS. J Diabetes Metab 2012;3(5):198-210.

(38) Kendler KS, Aggen SH, Patrick CJ. A multivariate twin study of the DSM-IV criteria for antisocial personality disorder. Biol Psychiatry 2012;71(3):247-253.

(39) Bergen SE, Petryshen TL. Genome-wide association studies of schizophrenia: does bigger lead to better results? Curr Opin Psychiatry 2012;25(2):76-82.

(40) Atsma F, Veldhuizen I, de Vegt F, Doggen C, de Kort W. Cardiovascular and demographic character istics in whole blood
and plasma donors: results from the Donor In Sight study. Transfusion 2011;51(2):412-420.

(41) Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet 2012 May;131(5):747-756.

(42) Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003 Aug 5;100(16):9440-9445.

(43) Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. Ann Stat 2003;31(6):2013-2035.

(44) Crabb DW, Matsumoto M, Chang D, You M. Overview of the role of alcohol dehydrogenase and aldehyde dehydrogenase and their variants in the genesis of alcohol-related pathology. Proc Nutr Soc 2004 Feb;63(1):49-63.

(45) Osier MV, Pakstis AJ, Soodyall H, Comas D, Goldman D, Odunsi A, et al. A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. Am J Hum Genet 2002 Jul;71(1):84-99.

(46) Frank J, Cichon S, Treutlein J, Mann K, Mann K, Nöthen M M, et al. Genome-wide significant association between alcohol
dependence and a variant in the ADH gene cluster. Addiction Biology 2011;17(1):171-180.

(47) Birley AJ, James MR, Dickson PA, Montgomery GW, Heath AC, Martin NG, et al. ADH single nucleotide polymorphism associations with alcohol metabolism in vivo. Hum Mol Genet 2009 Apr 15;18(8):1533-1542.

(48) Osier M, Pakstis AJ, Kidd JR, Lee JF, Yin SJ, Ko HC, et al. Linkage disequilibrium at the ADH2 and ADH3 loci and risk of alcoholism. Am J Hum Genet 1999 Apr;64(4):1147-1157.

(49) Peng GS, Yin SJ. Effect of the allelic variants of aldehyde dehydrogenase ALDH2*2 and alcohol dehydrogenase ADH1B*2 on blood acetaldehyde concentrations. Hum Genomics 2009 Jan;3(2):121-127.

(50) Edenberg HJ, Xuei X, Chen HJ, Tian H, Wetherill LF, Dick DM, et al. Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. Hum Mol Genet 2006 May 1;15(9):1539-1549.

(51) Luo X, Kranzler HR, Zuo L, Lappalainen J, Yang BZ, Gelernter J. ADH4 gene variation is associated with alcohol dependence and drug dependence in European Americans: results from HWD tests and case-control association studies. Neuropsychopharmacology 2006 May;31(5):1085-1095.

(52) Chai YG, Oh DY, Chung EK, Kim GS, Kim L, Lee YS, et al. Alcohol and aldehyde dehydrogenase polymorphisms in men with type I and Type II alcoholism. Am J Psychiatry 2005 May;162(5):1003-1005.

(53) Takeuchi F, Isono M, Nabika T, Katsuya T, Sugiyama T, Yamaguchi S, et al. Confirmation of ALDH2 as a Major Locus of Drinking Behavior and of Its Variants Regulating Multiple Metabolic Phenotypes in a Japanese Population. Circulation Journal 2011;75:911-918.

(54) Enoch MA. The role of GABA(A) receptors in the development of alcoholism. Pharmacol Biochem Behav 2008 Jul;90(1):95-104.

(55) Dick DM, Plunkett J, Wetherill LF, Xuei X, Goate A, Hesselbrock V, et al. Association between GABRA1 and drinking behaviors in the collaborative study on the genetics of alcoholism sample. Alcohol Clin Exp Res 2006 Jul;30(7):1101-1110.

(56) Spanagel R. Alcoholism: a systems approach from molecular physiology to addictive behavior. Physiol Rev 2009 Apr;89(2):649-705.

(57) Le Foll B, Gallo A, Le Strat Y, Lu L, Gorwood P. Genetics of dopamine receptors and drug addiction: a comprehensive review. Behav Pharmacol 2009;20(1):1-17.

(58) Neville MJ, Johnstone EC, Walton RT. Identification and characterization of ANKK1: a novel kinase gene closely linked to DRD2 on chromosome band 11q23.1. Hum Mutat 2004 Jun;23(6):540-545.

(59) Dick DM, Wang JC, Plunkett J, Aliev F, Hinrichs A, Bertelsen S, et al. Family-based association analyses of alcohol dependence phenotypes across DRD2 and neighboring gene ANKK1. Alcohol Clin Exp Res 2007;31:1645-1653.

(60) Batel P, Houchi H, Daoust M, Ramoz N, Naassila M, Gorwood P. A haplotype of the DRD1 gene is associated with alcohol dependence. Alcohol Clin Exp Res 2008;32:567-572.

(61) Kim DJ, Park BL, Yoon S, Lee HK, Joe KH, Cheon YH, et al. 5' UTR polymorphism of dopamine receptor D1 (DRD1) associated with severity and temperament of alcoholism. Biochem Biophys Res Commun 2007;357:1135-1141.

(62) McGeary J. The DRD4 exon 3 VNTR polymorphism and addiction-related phenotypes: A review. Pharmacology, Biochemistry and Behavior 2009;93:222-229.

(63) Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, et al. Genome-wide association study of alcohol dependence. Arch Gen Psychiatry 2009 Jul;66(7):773-784.

(64) Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, et al. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcohol Clin Exp Res 2010 May;34(5):840-852.

(65) Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, et al. A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A 2010 Mar 16;107(11):5082-5087.

(66) Lind PA, Macgregor S, Vink JM, Pergadia ML, Hansell NK, de Moor MH, et al. A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations. Twin Res Hum Genet 2010 Feb;13(1):10-29.

(67) Johnson C, Drgon T, Liu QR, Walther D, Edenberg H, Rice J, et al. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. Am J Med Genet B Neuropsychiatr Genet 2006 Dec 5;141B(8):844-853.

(68) Kendler KS, Kalsi G, Holmans PA, Sanders AR, Aggen SH, Dick DM, et al. Genomewide association analysis of symptoms of alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample. Alcohol Clin Exp Res 2011 May;35(5):963-975.

(69) Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, et al. A Quantitative-Trait Genome-Wide Association Study of Alcoholism Risk in the Community: Findings and Implications. Biol Psychiatry 2011;70(6):513-518.

(70) Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. Proc Natl Acad Sci U S A 2011 Apr 26;108(17):7119-7124.

(71) Baik I, Cho NH, Kim SH, Han BG, Shin C. Genome-wide association studies identify genetic loci related to alcohol consumption in Korean men. Am J Clin Nutr 2011 Apr;93(4):809-816.

(72) Treutlein J, Rietschel M. Genome-wide association studies of alcohol dependence and substance use disorders. Curr Psychiatry Rep 2011;13(2):147-155.

(73) Latvala A, Dick DM, Tuulio-Henriksson A, Suvisaari J, Viken RJ, Rose RJ, et al. Genetic correlation and gene-environment interaction between alcohol problems and educational level in young adulthood. J Stud Alcohol Drugs 2011 Mar;72(2):210-220.

(74) Crum RM, Juon HS, Green KM, Robertson J, Fothergill K, Ensminger M. Educational achievement and early school behavior as predictors of alcohol-use disorders: 35-year follow-up of the Woodlawn Study. J Stud Alcohol 2006 Jan;67(1):75-85.

(75) Crum RM, Helzer JE, Anthony JC. Level of education and alcohol abuse and dependence in adulthood: a further inquiry. Am J Public Health 1993;83(6):830-837.

(76) Huerta MC, Borgonovi F. Education, alcohol use and abuse among young adults in Britain. Soc Sci Med 2010 Jul;71(1):143-151.

(77) Staff J, Patrick ME, Loken E, Maggs JL. Teenage Alcohol Use and Educational Attainment. J Stud Alcohol Drugs 2008;69(6):848-858.

(78) Chatterji P. Does alcohol use during high school affect educational attainment?: Evidence from the National Education Longitudinal Study. Economics of Education Review 2006;25:482-497.

(79) Baker LA, Treloar SA, Reynolds CA, Heath AC, Martin NG. Genetics of educational attainment in Australian twins: sex differences and secular changes. Behav Genet 1996 Mar;26(2):89-102.

(80) Tambs K, Kendler KS, Reichborn-Kjennerud T, Aggen SH, Harris JR, Neale MC, et al. Genetic and environmental contributions to the relationship between education and anxiety disorders - a twin study. Acta Psychiatr Scand 2011 Nov 24.

(81) Heath AC, Berg K, Eaves LJ, Solaas MH, Corey LA, Sundet J, et al. Education policy and the heritability of educational attainment. Nature 1985 Apr 25-May 1;314(6013):734-736.

(82) Tsuang MT, Bar JL, Stone WS, Faraone SV. Gene-environment interactions in mental disorders. World Psychiatry 2004 Jun;3(2):73-83.

(83) Heath AC, Jardine R, Martin NG. Interactive effects of genotype and social environment on alcohol consumption in female twins. J Stud Alcohol 1989 Jan;50(1):38-48.

(84) Koopmans JR, Slutske WS, van Baal GC, Boomsma DI. The influence of religion on alcohol use initiation: Evidence for Genotype × Environment interaction. Behav Genet 1999;29(6):445-453.

(85) Dick DM, Rose RJ, Viken RJ, Kaprio J, Koskenvuo M. Exploring gene-environment interactions: socioregional moderation of alcohol use. J Abnorm Psychol 2001 Nov;110(4):625-632.

(86) Rose RJ, Dick DM, Viken And RJ, Kaprio J. Gene-environment interaction in patterns of adolescent drinking: regional residency moderates longitudinal influences on alcohol use. Alcohol Clin Exp Res 2001 May;25(5):637-643.

(87) Harden KP, Hill JE, Turkheimer E, Emery RE. Gene-environment correlation and interaction in peer effects on adolescent alcohol and tobacco use. Behav Genet 2008 Jul;38(4):339-347.

(88) Kendler KS, Gardner C, Dick DM. Predicting alcohol consumption in adolescence from alcohol-specific and general externalizing genetic risk factors, key environmental exposures and their interaction. Psychol Med 2011 Jul;41(7):1507-1516.

(89) Dick DM, Pagan JL, Viken R, Purcell S, Kaprio J, Pulkkinen L, et al. Changing environmental influences on substance use across development. Twin Res Hum Genet 2007 Apr;10(2):315-326.

(90) Dick DM, Viken R, Purcell S, Kaprio J, Pulkkinen L, Rose RJ. Parental monitoring moderates the importance of genetic and environmental influences on adolescent smoking. J Abnorm Psychol 2007 Feb;116(1):213-218.

(91) Kendler KS, Prescott CA. Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders. New York, NY: The Guildford Press; 2006.

(92) Spitzer RL, Williams JBW. Structured Clinical Interview for DSM-III-R (SCID). New York State Psychiatric Institute, Biometrics Research Department, New York; 1985.

(93) SAS Institute Inc. SAS 9.2 Language Reference: Concepts. 2010.

(94) Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet 2002 Nov;3(11):872-882.

(95) Verweij KJ, Mosing MA, Zietsch BP, Medland SE. Estimating heritability from twin studies. Methods Mol Biol 2012;850:151-170.

(96) Neale MC. Twin Studies: Software and Algorithms. Encyclopedia of Life Sciences London: Wiley & Sons Ltd.; 2001. p. 679-683.

(97) Neale M, Cardon L. Methodology for Genetic Studies of Twins and Families. Dordrecht, The Netherland: Kluwer Academic Publishers. B.V.; 1992.

(98) Posthuma D, Beem AL, de Geus EJ, van Baal GC, von Hjelmborg JB, Iachine I, et al. Theory and practice in quantitative genetics. Twin Res 2003 Oct;6(5):361-376.

(99) Akaike H. Factor analysis and AIC. Psychometrika 1987;52:317-332.

(100) Williams L, Holahan P. Parsimony-based fit indices for multiple-indicator models: do they work? Structural Equation Modeling 1994;1:161-189.

(101) Neale MC, Boker SM, Xie G, Maes HH. Mx: Statistical Modeling. Richmond, Virginia Commonwealth University Medical School, Department of Psychiatry; 2003.

(102) Kendler KS, Chen X, Dick D, Maes H, Gillespie N, Neale MC, et al. Recent advances in the genetic epidemiology and molecular genetics of substance use disorders. Nat Neurosci 2012 Jan 26;15(2):181-189.

(103) Purcell S. Variance components models for gene-environment interaction in twin analysis. Twin Res 2002 Dec;5(6):554-571.

(104) Latvala A, Tuulio-Henriksson A, Dick DM, Vuoksimaa E, Viken RJ, Suvisaari J, et al. Genetic origins of the association between verbal ability and alcohol dependence symptoms in young adulthood. Psychol Med 2011 Mar;41(3):641-651.

(105) Johnson W, Hicks BM, McGue M, Iacono WG. How Intelligence and Education Contribute to Substance Use: Hints from the Minnesota Twin Family Study. Intelligence 2009 Nov 1;37(6):613-624.

(106) Jefferis BJ, Manor O, Power C. Cognitive development in childhood and drinking behaviour over two decades in adulthood. J Epidemiol Community Health 2008 Jun;62(6):506-512.

(107) McCaffery JM, Papandonatos GD, Lyons MJ, Koenen KC, Tsuang MT, Niaura R. Educational attainment, smoking initiation and lifetime nicotine dependence among male Vietnam-era twins. Psychol Med 2008 Sep;38(9):1287-1297.

(108) Sen S. Socioeconomic status and mental health: what is the causal relationship?: Editorial comment to Kristian Tambs et al. 'Genetic and environmental contributions to the relationship between education and anxiety disorders. A twin study' (1). Acta Psychiatr Scand 2012 Mar;125(3):187-188.

(109) Russel H, O'Connell P. Getting a Job in Europe: The Transition from Unemployment to Work among Young People in Nine European Countries. Work Employment & Society 2001;15(1):1-24.

(110) Edwards AC, Kendler KS. Twin study of the relationship between adolescent attention-deficit/hyperactivity disorder and adult alcohol dependence. J Stud Alcohol Drugs 2012 Mar;73(2):185-194.

(111) Feinberg ME, Button TM, Neiderhiser JM, Reiss D, Hetherington EM. Parenting and adolescent antisocial behavior and depression: evidence of genotype x parenting environment interaction. Arch Gen Psychiatry 2007 Apr;64(4):457-465.

(112) Robbers S, van Oort F, Huizink A, Verhulst F, van Beijsterveldt C, Boomsma D, et al. Childhood problem behavior and parental divorce: evidence for gene-environment interaction. Soc Psychiatry Psychiatr Epidemiol 2012 Jan 13.

(113) Young-Wolff KC, Enoch MA, Prescott CA. The influence of gene-environment interactions on alcohol consumption and alcohol use disorders: a comprehensive review. Clin Psychol Rev 2011 Jul;31(5):800-816.

(114) Hicks BM, South SC, Dirago AC, Iacono WG, McGue M. Environmental Adversity Increases Genetic Risk for Externalizing Disorders. Hicks BM, South SC, Dirago AC, Iacono WG, McGue M 2009;66(6):640-648.

(115) Alcacera MA, Marques-Lopes I, Fajo-Pascual M, Foncillas JP, Carmona-Torre F, Martinez-Gonzalez MA. Alcoholic beverage preference and dietary pattern in Spanish university graduates: the SUN cohort study. Eur J Clin Nutr 2008 Oct;62(10):1178-1186.

(116) Becker U, Grønbaek M, Johansen D, Sørensen TI. Lower risk for alcohol-induced cirrhosis in wine drinkers. Hepatology 2002;35:868-875.

(117) Grønbaek M, Becker U, Johansen D, Gottschau A, Schnohr P, Hein HO, et al. Type of alcohol consumed and mortality from all causes, coronary heart disease, and cancer. Ann Intern Med 2000;133:411-419.

(118) Høidrup S, Grønbæk M, Gottschau A, Lauritzen JB, Schroll M. Alcohol intake, beverage preference, and risk of hip fracture in men and women. Copenhagen Centre for Prospective Population Studies. American Journal of Epidemiology 1999;149:993-1001.

(119) Petri AL, Tjønneland A, Gamborg M, Johansen D, Høidrup S, Sorensen TI, et al. Alcohol intake, type of beverage, and risk of breast cancer in pre- and post menopausal women. Alcoholism: Clinical and Experimental Research 2004;28:1084-1090.

(120) Flensborg-Madsen T, Knop J, Mortensen EL, Becker U, Makhija N, Sher L, et al. Beverage preference and risk of alcohol-use disorders: a Danish prospective cohort study. Journal of Studies on Alcohol and Drugs 2008;69:371-377.

(121) Gruenewald PJ, Johnson FW, Millar A, Mitchell PR. Drinking and driving: explaining beverage-specific risks. Journal of Studies on Alcohol and Drugs 2000;61:515-523.

(122) Jensen MK, Andersen AT, Sørensen TI, Becker U, Thorsen T, Grønbaek M. Alcoholic beverage preference and risk of becoming a heavy drinker. Epidemiology 2002;13:120-122.

(123) Lintonen TP, Konu AI. Adolescent Alcohol Beverage Type Choices Reflect TheirSubstance Use Patterns and Attitudes. Journal of Youth and Adolescence 2003;32:279-289.

(124) Smart RG, Walsh GW. Do some types of alcoholic beverages lead to more problems for adolescents? Journal of Studies on Alcohol 1995;56:35-38.

(125) Smart RG. Behavioral and social consequences related to the consumption of different beverage types. J Stud Alcohol 1996 Jan;57(1):77-84.

(126) Smart RG, Walsh GW. Heavy drinking and problems among wine drinkers. Journal of Studies on Alcohol 1999;60:467-471.

(127) Sayon-Orea C, Bes-Rastrollo M, Nuñez-Cordoba JM, Basterra-Gortari FJ, Beunza JJ, Martinez-Gonzalez MA. Type of alcoholic beverage and incidence of overweight/obesity in a Mediterranean cohort: the SUN project. Nutrition 2011;27:802-808.

(128) Prescott E, Grønbaek M, Becker U, Sørensen TI. Alcohol intake and the risk of lung cancer: influence of type of alcoholic beverage. Am J Epidemiol 1999;149:463-470.

(129) Truelsen T, Gronbaek M, Schnohr P, Boysen G. Intake of beer, wine, and spirits and risk of stroke: the copenhagen city heart study. Stroke 1998;29:2467-2472.

(130) Dufour MC. What is moderate drinking? Defining "drinks" and drinking levels. Alcohol Res Health 1999;23(1):5-14.

(131) Huang WY, Winn DM, Brown LM, Gridley G, Bravo-Otero E, Diehl SR, et al. Alcohol concentration and risk of oral cancer in Puerto Rico. Am J Epidemiol 2003 May 15;157(10):881-887.

(132) Yi SW, Sull JW, Linton JA, Nam CM, Ohrr H. Alcohol consumption and digestive cancer mortality in Koreans: the Kangwha Cohort Study. J Epidemiol 2010;20(3):204-211.

(133) Coder B, Freyer-Adam J, Lau K, Riedel J, Rumpf HJ, Meyer C, et al. Reported beverage consumed and alcohol-related diseases among male hospital inpatients with problem drinking. Alcohol Alcohol 2009 Mar-Apr;44(2):216-221.

(134) Wu JM, Hsieh TC. Resveratro l: a cardioprotective substance. Ann N Y Acad Sci 2011;1215:16-21.

(135) Lachenmeier DW, Lima MC, Nobrega IC, Pereira JA, Kerr-Correa F, Kanteres F, et al. Cancer risk assessment of ethyl carbamate in alcoholic beverages from Brazil with special consideration to the spirits cachaca and tiquira. BMC Cancer 2010 Jun 8;10:266.

(136) Lachenmeier DW, Kanteres F, Rehm J. Carcinogenicity of acetaldehyde in alcoholic beverages: risk assessment outside ethanol metabolism. Addiction 2009 Apr;104(4):533-550.

(137) Arranz S C, G., Valderas-Martínez P, Medina-Remón A, Lamuela-Raventós RM, Estruch R. Wine, beer, alcohol and polyphenols on cardiovascular disease and cancer**.**. Nutrients 2012;4:759-781.

(138) Kaprio J, Rose RJ, Romanov K, Koskenvuo M. Genetic and environmental determinants of use and abuse of alcohol: the Finnish Twin Cohort studies. Alcohol Alcohol Suppl 1991;1:131-136.

(139) Kaprio J, Koskenvuo M, Langinvainio H, Romanov K, Sarna S, Rose RJ. Genetic influences on use and abuse of alcohol: a study of 5638 adult Finnish twin brothers. Alcohol Clin Exp Res 1987 Aug;11(4):349-356.

(140) Fabsitz RR, Garrison RJ, Feinleib M, Hjortland M. A twin analysis of dietary intake: evidence for a need to control for possible environmental differences in MZ and DZ twins. Behavior Genetics 1978;8:15-25.

(141) Falconer DS. Introduction to Quantitative Genetics. New York: Ronald Press; 1960.

(142) Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JIJ, et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J of Stud Alcohol 1994;55(2):149-58.

(143) Kaprio J. Twin studies in Finland 2006. Twin Res Hum Genet 2006 Dec;9(6):772-777.

(144) Viini Web. 2010; Available at: http://www.viiniweb.net/winefact.php. Accessed 12/05, 2010.

(145) Osterberg E. Finland. In: Blocker JS, Fahey DM, Tyrrell IR, editors. Alcohol and Temperance in Modern History: An International Encyclopedia. Library of Congress Cataloging-in-Publication ed.; 2003. p. 240-242.

(146) Luciano M, Kirk KM, Heath AC, Martin NG. The genetics of tea and coffee drinking and preference for source of caffeine in a large community sample of Australian twins. Addiction 2005 Oct;100(10):1510-1517.

(147) Vink JM, Staphorsius AS, Boomsma DI. A genetic analysis of coffee consumption in a sample of Dutch twins. Twin Res Hum Genet 2009 Apr;12(2):127-131.

(148) Truett KR, Eaves LJ, Walters EE, Heath AC, Hewitt JK, Meyer JM, et al. A model system for analysis of family resemblance in extended kinships of twins. Behav Genet 1994 Jan;24(1):35-49.

(149) Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, et al. OpenMx: An Open Source Extended Structural Equation Modeling Framework. Psychometrika 2011;76:306-317.

(150) Vinkhuyzen AA, van der Sluis S, Maes HH, Posthuma D. Reconsidering the Heritability of Intelligence in Adulthood: Taking Assortative Mating and Cultural Transmission into Account. Behav Genet 2011 Oct 4.

(151) Heath AC, Eaves LJ. Resolving the effects of phenotype and social background on mate selection. Behav Genet 1985 Jan;15(1):15-30.

(152) Keller MC, Medland SE, Duncan LE, Hatemi PK, Neale MC, Maes HH, et al. Modeling extended twin family data I: description of the Cascade model. Twin Res Hum Genet 2009 Feb;12(1):8-18.

(153) Maes HH, Neale MC, Medland SE, Keller MC, Martin NG, Heath AC, et al. Flexible Mx specification of various extended twin kinship designs. Twin Res Hum Genet 2009 Feb;12(1):26-34.

(154) Hansell S, White HR, Vali FM. Specific Alcoholic Beverages and Physical and Mental Health among Adolescents. Journal of Studies on Alcohol 1999;60:209-218.

(155) Klatsky AL, Armstrong MA, Kipp H. Correlates of alcoholic beverage preference: Traits of persons who choose wine, liquor, or beer. British Journal of Addiction 1990;85:1279-1289.

(156) Marques-Vidal P, Dias CM. Trends and determinants of alcohol consumption in Portugal: results from the national health surveys 1995 to 1996 and 1998 to 1999. Alcohol Clin Exp Res 2005;29:89-97.

(157) Rogers JD, Greenfield TK. Beer drinking accounts for most of the hazardous alcohol consumption reported in the United States. Journal of Studies on Alcohol 1995;56:35-38.

(158) Naimi TS, Brewer RD, Miller JW, Okoro C, Mehrotra C. What do binge drinkers drink? Implications for alcohol control policy. Am J Prev Med 2007 Sep;33(3):188-193.

(159) Berger DE, Snortum JR. Alcoholic beverage preferences of drinking-driving violators. J Stud Alcohol 1985 May;46(3):232-239.

(160) Greenfield TK, Rogers JD. Alcoholic beverage choice, risk perception and self-reported drunk driving: effects of measurement on risk analysis. Addiction 1999;94:1735-1743.

(161) Siegel MB, Naimi TS, Cremeens JL, Nelson DE. Alcoholic Beverage Preferences and Associated Drinking Patterns and Risk Behaviors Among High School Youth. American Journal of Prevention Medicine 2011;40:419-426.

(162) Caetano R, Vaeth PA, Ramisetty-Mikler S, Rodriguez LA. The Hispanic Americans Baseline Alcohol Survey (HABLAS): Alcoholic beverage preference across Hispanic national groups. Alcohol Clin Exp Res 2009;33(1):150-159.

(163) Graves K, Kaskutas LA. Beverage choice among native american and african american urban women. Alcohol Clin Exp Res 2002 Feb;26(2):218-222.

(164) Baltieri DA, Daro FR, Ribeiro PL, De Andrade AG. The role of alcoholic beverage preference in the severity of alcohol dependence and adherence to the treatment. Alcohol 2009 May;43(3):185-195.

(165) Norstrom T. Effects on criminal violence of different beverage types and private and public drinking. Addiction 1998 May;93(5):689-699.

(166) Mortensen EL, Jensen HH, Sanders SA, Reinisch JM. Better psychological functioning and higher social status may largely explain the apparent health benefits of wine: a study of wine and beer drinking in young Danish adults. Arch Intern Med 2001;161:1844-1848.

(167) Hesselbrock VM, Hesselbrock MN. Are there empirically supported and clinically useful subtypes of alcohol dependence? Addiction 2006 Sep;101 Suppl 1:97-103.

(168) Wetterling T, Veltrup C, John U, Driessen M. Late onset alcoholism. Eur Psychiatry 2003 May;18(3):112-118.

(169) Klein H, Pittman DJ. Alcohol consumption and the perceived situational appropriateness of consuming different types of alcoholic beverages. Alcohol Alcohol 1989;24(5):479-488.

(170) Ishibashi T, Harada S, Fujii C, Taguchi A, Ishii T. Relationship between ALDH2 genotypes and choice of alcoholic beverages. Nihon Arukoru Yakubutsu Igakkai Zasshi 1999 Apr;34(2):117-129.

(171) Hendershot CS, Collins SE, George WH, Wall TL, McCarthy DM, Liang T, et al. Associations of ALDH2 and ADH1B genotypes with alcohol-related phenotypes in Asian young adults. Alcohol Clin Exp Res 2009 May;33(5):839-847.

(172) Duffy VB, Davidson AC, Kidd JR, Kidd KK, Speed WC, Pakstis AJ, et al. Bitter receptor gene (TAS2R38), 6-n-propylthiouracil (PROP) bitterness and alcohol intake. Alcoholism: Clinical and Experimental Research 2004;28(1629):1637.

(173) Lanier SA, Hayes JE, Duffy VB. Sweet and bitter tastes of alcoholic beverages medicate alcohol intake in of-age undergraduates. Physiology & Behavior 2005;83:821-831.

(174) Hinrichs AL, Wang JC, Bufe B, Kwon JM, Budde J, Allen R, et al. Functional variant in bitter-taste receptor (hTASR16) influences risk of alcohol dependence. American Journal of Human Genetics 2006;78:103-111.

(175) Hillemacher T, Bayerlein K, Reulbach U, Sperling W, Wilhelm J, Mugele B, et al. Influence of beer, wine and spirits consumption on craving. Addict Biol 2005;10(2):181-186.

(176) Mennen LI, de Courcy GP, Guilland JC, Ducros V, Zarebska M, Bertrais S, et al. Relation between homocysteine concentrations and the consumption of different types of alcoholic beverages: the French Supplementation with Antioxidant Vitamins and Minerals Study. Am J Clin Nutr 2003 Aug;78(2):334-338.

(177) Lutz UC. Alterations in homocysteine metabolism among alcohol dependent patients--clinical, pathobiochemical and genetic aspects. Curr Drug Abuse Rev 2008 Jan;1(1):47-55.

(178) Wilhelm J, Frieling H, Hillemacher T, Degner D, Kornhuber J, Bleich S. Hippocampal volume loss in patients with alcoholism is influenced by the consumed type of alcoholic beverage. Alcohol Alcohol 2008;43:296-299.

(179) Truelsen T, Thudium D, Grønbaek M. Amount and type of alcohol and risk of dementia: the Copenhagen City Heart Study. Neurology 2002;59:1313-1319.

(180) Rohsenow DJ, Howland J, Arnedt JT, Almeida AB, Greece J, Minsky S, et al. Intoxication with bourbon versus vodka: effects on hangover, sleep, and next-day neurocognitive performance in young adults. Alcohol Clin Exp Res 2010 Mar 1;34(3):509-518.

(181) Rohsenow DJ, Howland J. The role of beverage congeners in hangover and other residual effects of alcohol intoxication: a review. Curr Drug Abuse Rev 2010 Jun;3(2):76-79.

(182) Klatsky AL, Friedman GD, Armstrong MA, Kipp H. Wine, liquor, beer, and mortality. Am J Epidemiol 2003 Sep 15;158(6):585-595.

(183) Barefoot JC, Gronbaek M, Feaganes JR, McPherson RS, Williams RB, Siegler IC. Alcoholic beverage preference, diet, and health habits in the UNC Alumni Heart Study. Am J Clin Nutr 2002 Aug;76(2):466-472.

(184) Graziano F, Bina M, Giannotta F, Ciairano S. Drinking motives and alcoholic beverage preferences among Italian adolescents. J Adolesc 2012 Jan 10.

(185) Fanous AH, Kendler KS. Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: searching for a framework. Mol Psychiatry 2005 Jan;10(1):6-13.

(186) Panconesi A. Alcohol and migraine: trigger factor, consumption, mechanisms. A review. J Headache Pain 2008 Feb;9(1):19-27.

(187) Prescott CA, Aggen SH, Kendler KS. Sex differences in the sources of genetic liability to alcohol abuse and dependence in a population-based sample of U.S. twins. Alcohol Clin Exp Res 1999 Jul;23(7):1136-1144.

(188) Walters GD. The heritability of alcohol abuse and dependence: A meta-analysis of behavior genetic research. American Journal of Alcohol Abuse 2002;28:557-584.

(189) Hardie TL, Moss HB, Lynch KG. Sex difference in the heritability of alcohol problems. Am J Addict 2008;17:319-327.

(190) de Castro JM. A twin study of genetic and environmental influences on the intake of fluids and beverages. Physiol Behav 1993 Oct;54(4):677-687.

(191) Hasselbalch AL, Silventoinen K, Keskitalo K, Pietilainen KH, Rissanen A, Heitmann BL, et al. Twin study of heritability of eating bread in Danish and Finnish men and women. Twin Res Hum Genet 2010 Apr;13(2):163-167.

(192) Borsari B, Carey KB. Peer influences on college drinking: a review of the research. J Subst Abuse 2001;13(4):391-424.

(193) Landrine H, Bardwell S, Dean T. Gender expectations for alcohol use: A study of the significance of the masculine role. Sex Roles 1988;19:703-712.

(194) Corcoran KJ, Segrist DJ. Fear of negative evaluation and gender interact to predict alcoholic beverage preference. Addict Behav 1998 Jul-Aug;23(4):509-515.

(195) Bartoshuk LM, Duffy VB, Miller IJ. PTC/PROP tasting: anatomy, psychophysics, and sex effects. Physiol Behav 1994 Dec;56(6):1165-1171.

(196) Driscoll KA, Perez M, Cukrowicz KC, Butler M, Joiner TE,Jr. Associations of phenylthiocarbamide tasting to alcohol problems and family history of alcoholism differ by gender. Psychiatry Res 2006 Jun 30;143(1):21-27.

(197) Kerr WC, Greenfield TK, Bond J, Ye Y, Rehm J. Age, period and cohort influences on beer, wine and spirits consumption trends in the US National Alcohol Surveys. Addiction 2004 Sep;99(9):1111-1120.

(198) Maes HH, Neale MC, Kendler KS, Martin NG, Heath AC, Eaves LJ. Genetic and cultural transmission of smoking initiation: an extended twin kinship model. Behav Genet 2006 Nov;36(6):795-808.

(199) (200) Strandberg TE, Strandberg AY, Salomaa VV, Pitkälä K, Tilvis RS, Miettinen TA. Alcoholic beverage preference, 29-year mortality, and quality of life in men in old age. J Gerontol A Biol Sci Med Sci 2007;62:213-218.

(201) Kubo A, Levin TR, Block G, Rumore GJ, Quesenberry CP,Jr, Buffler P, et al. Alcohol types and sociodemographic characteristics as risk factors for Barrett's esophagus. Gastroenterology 2009 Mar;136(3):806-815.

(202) Mortensen LH, Sørensen TI, Grønbaek M. Intelligence in relation to later beverage preference and alcohol intake. Addiction 2005;100:1445-1452.

(203) Tjønneland A, Grønbaek M, Stripp C, Overvad K. Wine intake and diet in a random sample of 48763 Danish men and women. Am J Clin Nutr 1999;69:49-54.

(204) Volcik KA, Ballantyne CM, Fuchs FD, Sharrett AR, Boerwinkle E. Relationship of alcohol consumption and type of alcoholic beverage consumed with plasma lipid levels: differences between Whites and African Americans of the ARIC study. Ann Epidemiol 2008 Feb;18(2):101-107.

(205) Kuntsche E, Knibbe R, Gmel G, Engels R. 'I drink spirits to get drunk and block out my problems...' beverage preference, drinking motives and alcohol use in adolescence. Alcohol Alcohol 2006 Sep-Oct;41(5):566-573.

(206) Agrawal A, Prescott CA, Kendler KS. Forms of cannabis and cocaine: a twin study. Am J Med Genet B Neuropsychiatr Genet 2004 Aug 15;129B(1):125-128.

(207) Derringer J, Krueger RF, Dick DM, Saccone S, Grucza RA, Agrawal A, et al. Predicting sensation seeking from dopamine genes. A candidate-system approach. Psychol Sci 2010 Sep;21(9):1282-1290.

(208) Koob GF. Drugs of abuse: anatomy, pharmacology and function of reward pathways. Trends Pharmacol Sci 1992;13:177-184.

(209) Olds J, Milner P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. J Comp Physiol Psychol 1954;47:419-427.

(210) Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol 1998;80:1-27.

(211) Sintov ND, Kendler KS, Young-Wolff KC, Walsh D, Patterson DG, Prescott CA. Empirically defined subtypes of alcohol dependence in an Irish family sample. Drug Alcohol Depend 2010 Mar 1;107(2-3):230-236.

(212) Zucker R. Alcohol use and alcohol use disorders: A developmental-biopsychosocial systems formulation covering the life course. In: Cicchetti D, Cohen DJ, editors. Developmental Psychopathology. 2nd ed. ed. New York: Wiley; 2006.

(213) Jang KL, Vernon PA, Livesley WJ. Personality disorder traits, family environment, and alcohol misuse: A multivariate behavioral genetic analysis. Addition 2000;95:873-888.

(214) Young SE, Stallings MC, Corley RP, Krauter KS, Hewitt JK. Genetic and environmental influences on behavioral disinhibition. Am J Med Genet 2000;96:684-695.

(215) Dick DM, Viken RJ, Kaprio J, Pulkkinen L, Rose RJ. Understanding the covariation among childhood externalizing symptoms: genetic and environmental influences on conduct disorder, attention deficit hyperactivity disorder, and oppositional defiant disorder symptoms. J Abnorm Child Psychol 2005 Apr;33(2):219-229.

(216) Eaves L, Rutter M, Silberg JL, Shillady L, Maes H, Pickles A. Genetic and environmental causes of covariation in interview assessments of disruptive behavior in child and adolescent twins. Behav Genet 2000;30:321-334.

(217) Knopik VS, Heath AC, Bucholz KK, Madden PA, Waldron M. Genetic and environmental influences on externalizing behavior and alcohol problems in adolescence: a female twin study. Pharmacol Biochem Behav 2009 Sep;93(3):313-321.

(218) Nadder TS, Silberg JL, Eaves LJ, Maes HH, Meyer JM. Genetic effects on ADHD symptomatology in 7- to 13-year-old twins: results from a telephone survey. Behav Genet 1998 Mar;28(2):83-99.

(219) Silberg J, Rutter M, Meyer J, Maes H, Hewitt J, Simonoff E, et al. Genetic and environmental influences on the covariation between hyperactivity and conduct disturbance in juvenile twins. J Child Psychol Psychiatry 1996 Oct;37(7):803-816.

(220) Tuvblad C, Zheng M, Raine A, Baker LA. A common genetic factor explains the covariation among ADHD ODD and CD symptoms in 9-10 year old boys and girls. J Abnorm Child Psychol 2009 Feb;37(2):153-167.

(221) Gizer IR, Ficks C, Waldman ID. Candidate gene studies of ADHD: a meta-analytic review. Hum Genet 2009 Jul;126(1):51-90.

(222) Munafò MR, Yalcin B, Willis-Owen SA, Flint J. Association of the dopamine D4 receptor (DRD4) gene and approach-related personality traits: meta-analysis and new data. Biol Psychiatry 2008;63:197-206.

(223) Schwab SG, Franke PE, Hoefgen B, Guttenthaler V, Lichtermann D, Trixler M, et al. Association of DNA polymorphisms in the synaptic vesicular amine transporter gene (SLC18A2) with alcohol and nicotine dependence. Neuropsychopharmacology 2005;30:2263-2268.

(224) Lind PA, Eriksson CJ, Wilhelmsen KC. Association between harmful alcohol consumption behavior and dopamine transporter (DAT1) gene polymorphisms in a male Finnish population. Alcohol Clin Exp Res 2009;29:417-429.

(225) van der Zwaluw CS, Engels RC, Buitelaar J, Verkes RJ, Franke B, Scholte RH. Polymorphisms in the dopamine transporter gene (SLC6A3/DAT1) and alcohol dependence in humans: a systematic review. Pharmacogenomics 2009;10:853-866.

(226) Bousman CA, Glatt SJ, Everall IP, Tsuang MT. Genetic association studies of methamphetamine use disorders: A systematic review and synthesis. Am J Med Genet B Neuropsychiatr Genet 2009 Dec 5;150B(8):1025-1049.

(227) Li Y, Shao C, Zhang D, Zhao M, Lin L, Yan P, et al. The effect of dopamine D2, D5 receptor and transporter (SLC6A3) polymorphisms on the cue-elicited heroin craving in Chinese. Am J Med Genet B Neuropsychiatr Genet 2006;141B:269-273.

(228) Guindalini C, Howard M, Haddley K, Laranjeira R, Collier D, Ammar N, et al. A dopamine transporter gene functional variant associated with cocaine abuse in a Brazilian sample. Proc Natl Acad Sci U S A 2006;103:4552-4557.

(229) Burt SA, Mikolajewski AJ. Preliminary evidence that specific candidate genes are associated with adolescent-onset antisocial behavior. Aggress Behav 2008;34:437-445.

(230) Guo G, Roettger ME, Shih JC. Contributions of the DAT1 and DRD2 genes to serious and violent delinquency among adolescents and young adults. Hum Genet 2007;121:125-136.

(231) Jorm AF, Prior M, Sanson A, Smart D, Zhang Y, Easteal S. Association of a polymorphism of the dopamine transporter gene with externalizing behavior problems and associated temperament traits: a longitudinal study from infancy to the mid-teens. Am J Med Genet 2001;105:346-350.

(232) Schulz-Heik RJ, Maentz SK, Rhee SH, Gelhorn HL, Young SE, Timberlake DS, et al. Case-control and within-family tests for an association between conduct disorder and DAT1. Psychiatr Genet 2008;18:17-24.

(233) Young SE, Smolen A, Corley RP, Krauter KS, DeFries JC, Crowley TJ, et al. Dopamine transporter polymorphism associated with externalizing behavior in children. Am J Med Genet B Neuropsychiatr Genet 2002;114:144-149.

(234) Dahmen N, Völp M, Singer P, Hiemke C, Szegedi A. Tyrosine hydroxylase Val-81-Met polymorphism associated with early-onset alcoholism. Psychiatr Genet 2005;15:13-16.

(235) Faraone SV, Khan SA. Candidate gene studies of attention-deficit/hyperactivity disorder. J Clin Psychiatry 2006;67:13-20.

(236) Lachman HM, Papolos DF, Saito T, Yu YM, Szumlanski CL, Weinshilboum RM. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. Pharmacogenetics 1996;6:243-250.

(237) Köhnke MD. Approach to the genetics of alcoholism: A review based on pathophysiology. Biochem Pharmacol 2008;75:160-177.

(238) Golimbet VE, Alfimova MV, Gritsenko IK, Ebstein RP. Relationship between dopamine system genes and extraversion and novelty seeking. Neurosci Behav Physiol 2007;37(6):601-606.

(239) Tsai SJ, Hong CJ, Yu YW, Chen TJ. Association study of catechol-O-methyltransferase gene and dopamine D4 receptor gene polymorphisms and personality traits in healthy young Chinese females. Neuropsychobiology 2004;50:153-156.

(240) Cheuk DK, Wong V. Meta-analysis of association between a catechol-O-methyltransferase gene polymorphism and attention deficit hyperactivity disorder. Behav Genet 2006;36(5):651-659.

(241) McCourt WF, Gurrera RJ, Cutter HS. Sensation seeking and novelty seeking. Are they the same? J Nerv Ment Dis 1993 May;181(5):309-312.

(242) Prescott CA, Sullivan PF, Myers JM, Patterson DG, Devitt M, Halberstadt LJ, et al. The Irish Affected Sib Pair Study of Alcohol Dependence: study methodology and validation of diagnosis by interview and family history. Alcohol Clin Exp Res 2005;29:417-429.

(243) Kuo PH, Neale MC, Riley BP, Webb BT, Sullivan PF, Vittum J, et al. Identification of susceptibility loci for alcohol-related traits in the Irish Affected Sib Pair Study of Alcohol Dependence. Alcohol Clin Exp Res 2006;30(11):1807-1816.

(244) Schuckit MA, Smith TL, Tipp JE. The Self-Rating of the Effects of alcohol (SRE) form as a retrospective measure of the risk for alcoholism. Addiction 1997 Aug;92(8):979-988.

(245) Wender PH. Minimal Brain Dysfunction In Children. New York: John Wiley & Sons, Inc.; 1971.

(246) Cloninger CR, Przybeck TR, Svrakic DM, Wetzel RD. The Temperament and Character Inventory (TCI): A Guide to Its Development and Use. Washington University, St. Louis: Center for Psychobiology of Personality; 1994.

(247) Hettema JM, An SS, Bukszar J, van den Oord EJ, Neale MC, Kendler KS, et al. Catechol-O-methyltransferase contributes to genetic susceptibility shared among anxiety spectrum phenotypes. Biol Psychiatry 2008;64:302-310.

(248) Hodgkinson CA, Yuan Q, Xu K, Shen PH, Heinz E, Lobos EA, et al. Addictions biology: haplotype-based analysis for 130 candidate genes on a single array. Alcohol Alcohol 2008;43:505-515.

(249) de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-1223.

(250) Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-265.

(251) Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007 Sep;81(3):559-575.

(252) R Development Core Team. R: A Language and Environment for Statistical Computing. 2011; Available at: http://www.R-project.org.

(253) Hakstian RA, Rogers WT, Cattell RB. The Behavior of Number-of-Factors Rules with Simulated Data. Multivariate Behavioral Research 1982;17(2):193-219.

(254) Gorwood P, Limosin F, Batel P, Duaux E, Gouya L, Adès J. The genetics of addiction: alcohol-dependence and D3 dopamine receptor gene. Pathol Biol (Paris) 2001;49:710-717.

(255) Lee MS, Ryu SH. No association between the dopamine D3 receptor gene and Korean alcohol dependence. Psychiatr Genet 2002;12:173-176.

(256) Wiesbeck GA, Dürsteler-MacFarland KM, Wurst FM, Walter M, Petitjean S, Müller S, et al. No association of dopamine receptor sensitivity in vivo with genetic predisposition for alcoholism and DRD2/DRD3 gene polymorphisms in alcohol dependence. Addict Biol 2006;11:72-75.

(257) Kaprio J. Genetic epidemiology of smoking behavior and nicotine dependence. 2009;6(4):304-306.

(258) Genome Wide Association Study of Schizophrenia in Ireland. Available at: http://www.wtccc.org.uk/ccc2/projects/ccc2_sp.shtml. Accessed 7/23, 2012.

(259) Irish Schizophrenia Genomics Consortium & the Wellcome Trust Case Control Consortium 2. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the MHC locus in schizophrenia. Biological Psychiatry 2012;In press.

(260) Affymetrix Power Tools. 2012; Available at: http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx. Accessed 9/15, 2011.

(261) The Broad Institute. **Birdsuite: Birdseed** . 2010; Available at: http://www.broadinstitute.org/mpg/birdsuite/birdseed.html. Accessed 6/15, 2011.

(262) Browning BL, Yu Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. Am J Hum Genet 2009 Dec;85(6):847-861.

(263) Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009 Jun;5(6):e1000529.

(264) Howie BN, Marchini J. IMPUTE2. 2012; Available at: http://mathgen.stats.ox.ac.uk/impute/impute_v2.html. Accessed 5/15, 2012.

(265) Thornton T, McPeek MS. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. Am J Hum Genet 2007 Aug;81(2):321-337.

(266) Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007;23(10):1294-1296.

(267) Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet 2010 Jul 9;87(1):139-145.

(268) van den Oord EJ, Sullivan PF. False discoveries and models for gene discovery. Trends Genet 2003 Oct;19(10):537-542.

(269) Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011 Sep 18;43(10):969-976.

(270) Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190-2191.

(271) Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. Genet Epidemiol 2010;34(1):60-66.

(272) Begum F, Ghosh D, Tseng GC, Feingold E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. Nucleic Acids Res 2012 May;40(9):3777-3784.

(273) Saccone SF, Quan J, Jones PL. BioQ: tracing experimental origins in public genomic databases using a novel data provenance model. Bioinformatics 2012 Apr 15;28(8):1189-1191.

(274) Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010 Sep 15;26(18):2336-2337.

(275) Prescott CA, Sullivan PF, Kuo PH, Webb BT, Vittum J, Patterson DG, et al. Genomewide linkage study in the Irish affected sib pair study of alcohol dependence: evidence for a susceptibility region for symptoms of alcohol dependence on chromosome 4. Mol Psychiatry 2006 Jun;11(6):603-611.

(276) Lampe AK, Bushby KM. Collagen VI related muscle disorders. J Med Genet 2005 Sep;42(9):673-685.

(277) Fox MA. Novel roles for collagens in wiring the vertebrate nervous system. Curr Opin Cell Biol 2008 Oct;20(5):508-513.

(278) Mas VR, Fassnacht R, Archer KJ, Maluf D. Molecular mechanisms involved in the interaction effects of alcohol and hepatitis C virus in liver cirrhosis. Mol Med 2010 Jul-Aug;16(7-8):287-297.

(279) Smith MJ, Culhane AC, Donovan M, Coffey JC, Barry BD, Kelly MA, et al. Analysis of differential gene expression in colorectal cancer and stroma using fluorescence-activated cell sorting purification. Br J Cancer 2009 May 5;100(9):1452-1464.

(280) Degen M, Brellier F, Schenk S, Driscoll R, Zaman K, Stupp R, et al. Tenascin-W, a new marker of cancer stroma, is elevated in sera of colon and breast cancer patients. Int J Cancer 2008 Jun 1;122(11):2454-2461.

(281) Scherberich A, Tucker R P, Degen M, Brown-Luedi M, Andres A C, Chiquet-Ehrismann R. Tenascin-W is found in malignant mammary tumors, promotes alpha8 integrin-dependent motility and requires p38MAPK activity for BMP-2 and TNF-alpha induced expression in vitro. Oncogene 2005;24(9):1525-1532.

(282) Gerjevic LN, Liu N, Lu S, Harrison-Findik DD. Alcohol Activates TGF-Beta but Inhibits BMP Receptor-Mediated Smad Signaling and Smad4 Binding to Hepcidin Promoter in the Liver. Int J Hepatol 2012;2012:459278.

(283) Bhandari P, Hill JS, Farris SP, Costin B, Martin I, Chan CL, et al. Chloride intracellular channels modulate acute ethanol behaviors in Drosophila, Caenorhabditis elegans and mice. Genes Brain Behav 2012 Jun;11(4):387-397.

(284) Zuo L, Gelernter J, Zhang CK, Zhao H, Lu L, Kranzler HR, et al. Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. Neuropsychopharmacology 2012 Jan;37(2):557-566.

(285) Giannini G, Conti A, Mammarella S, Scrobogna M, Sorrentino V. The ryanodine receptor/calcium channel genes are widely and differentially expressed in murine brain and peripheral tissues. J Cell Biol 1995 Mar;128(5):893-904.

(286) Chavis P, Ango F, Michel JM, Bockaert J, Fagni L. Modulation of big K+ channel activity by ryanodine receptors and L-type Ca2+ channels in neurons. Eur J Neurosci 1998 Jul;10(7):2322-2327.

(287) Davies AG, Pierce-Shimomura JT, Kim H, VanHoven MK, Thiele TR, Bonci A, et al. A central role of the BK potassium channel in behavioral responses to ethanol in C. elegans. Cell 2003;115(6):655-66.

(288) Blank MC, Grinberg I, Aryee E, Laliberte C, Chizhikov VV, Henkelman RM, et al. Multiple developmental programs are altered by loss of Zic1 and Zic4 to cause Dandy-Walker malformation cerebellar pathogenesis. Development 2011 Mar;138(6):1207-1216.

(289) Tonkin ET, Smith M, Eichhorn P, Jones S, Imamwerdi B, Lindsay S, et al. A giant novel gene undergoing extensive alternative splicing is severed by a Cornelia de Lange-associated translocation breakpoint at 3q26.3. Hum Genet 2004 Jul;115(2):139-148.

(290) Yanagisawa A, Bouchet C, Van den Bergh PY, Cuisset JM, Viollet L, Leturcq F, et al. New POMT2 mutations causing congenital muscular dystrophy: identification of a founder mutation. Neurology 2007 Sep 18;69(12):1254-1260.

(291) Van Houdt JK, Nowakowska BA, Sousa SB, van Schaik BD, Seuntjens E, Avonce N, et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. Nat Genet 2012 Feb 26;44(4):445-9, S1.

(292) Kutscher S, Heise DJ, Banger M, Saller B, Michel MC, Gastpar M, et al. Concomitant endocrine and immune alterations during alcohol intoxication and acute withdrawal in alcohol-dependent subjects. Neuropsychobiology 2002;45(3):144-149.

(293) Wilhelm J, Heberlein A, Karagulle D, Groschl M, Kornhuber J, Riera R, et al. Prolactin serum levels during alcohol withdrawal are associated with the severity of alcohol dependence and withdrawal symptoms. Alcohol Clin Exp Res 2011 Feb;35(2):235-239.

(294) Schuckit MA, Gold E, Risch C. Serum prolactin levels in sons of alcoholics and control subjects. Am J Psychiatry 1987 Jul;144(7):854-859.

(295) Ben-Jonathan N, Hnasko R. Dopamine as a prolactin (PRL) inhibitor. Endocr Rev 2001;22:724-763.

(296) Bregonzio C, Moreno GN, Cabrera RJ, Donoso AO. NMDA receptors in the medial zona incerta stimulate luteinizing hormone and prolactin release. Cell Mol Neurobiol 2004 Jun;24(3):331-342.

(297) Riley BP, Kalsi G, Kuo PH, Vladimirov V, Thiselton DL, Vittum J, et al. Alcohol dependence is associated with the ZNF699 gene, a human locus related to Drosophila hangover, in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) sample. Mol Psychiatry 2006 Nov;11(11):1025-1031.

(298) Havik B, Le Hellard S, Rietschel M, Lybaek H, Djurovic S, Mattheisen M, et al. The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia. Biol Psychiatry 2011 Jul 1;70(1):35-42.

(299) Lydall GJ, Bass NJ, McQuillin A, Lawrence J, Anjorin A, Kandaswamy R, et al. Confirmation of prior evidence of genetic susceptibility to alcoholism in a genome wide association study of comorbid alcoholism and bipolar disorder. Psychiatr Genet 2012;21(6):294-306.

(300) Wirgenes KV, Sonderby IE, Haukvik UK, Mattingsdal M, Tesli M, Athanasiu L, et al. TCF4 sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders. Transl Psychiatry 2012 May 8;2:e112.

(301) Araki K, Meguro H, Kushiya E, Takayama C, Inoue Y, Mishina M. Selective expression of the glutamate receptor channel delta 2 subunit in cerebellar Purkinje cells. Biochem Biophys Res Commun 1993 Dec 30;197(3):1267-1276.

(302) Agrawal A, Madden PA, Heath AC, Lynskey MT, Bucholz KK, Martin NG. Correlates of regular cigarette smoking in a population-based sample of Australian twins. Addiction 2005;100(11):1709-1719.

(303) Wang X, Prins BP, Sober S, Laan M, Snieder H. Beyond genome-wide association studies: new strategies for identifying genetic determinants of hypertension. Curr Hypertens Rep 2011 Dec;13(6):442-451.

(304) Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009 Jun 9;106(23):9362-9367.

(305) Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet 2011 Jun;43(6):513-518.

(306) Morozova TV, Goldman D, Mackay TF, Anholt RR. The genetic basis of alcoholism: multiple phenotypes, many genes, complex networks. Genome Biol 2012 Feb 20;13(2):239.

(307) Scholz H, Mustard JA. Invertebrate Models of Alcoholism. Curr Top Behav Neurosci 2011 Apr 7.

(308) McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010 Apr 16;141(2):210-217.

(309) Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009 Oct 8;461(7265):747-753.

(310) Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet 2012 Jan 18;13(2):135-145.

(311) International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009 Aug 6;460(7256):748-752.

(312) Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 2012 Mar 16;148(6):1223-1241.

(313) Lin P, Hartz SM, Wang JC, Agrawal A, Zhang TX, McKenna N, et al. Copy Number Variations in 6q14.1 and 5q13.2 are Associated with Alcohol Dependence. Alcohol Clin Exp Res 2012 Jun 15.

(314) Liu J, Calhoun VD, Chen J, Claus ED, Hutchison KE. Effect of homozygous deletions at 22q13.1 on alcohol dependence severity and cue-elicited BOLD response in the precuneus. Addict Biol 2011 Oct 13.

(315) Bae JS, Jung MH, Lee BC, Cheong HS, Park BL, Kim LH, et al. The genetic effect of copy number variations on the risk of alcoholism in a Korean population. Alcohol Clin Exp Res 2012;36(1):35-42.

(316) Pei YF, Zhang L, Yang TL, Han Y, Hai R, Ran S, et al. Genome-wide association study of copy number variants suggests LTBP1 and FGD4 are important for alcohol drinking. PLoS One 2012;7(1):e30860.

(317) Boutte D, Calhoun VD, Chen J, Sabbineni A, Hutchison K, Liu J. Association of genetic copy number variations at 11 q14.2 with brain regional volume differences in an alcohol use disorder population. Alcohol 2012 Jun 23.

(318) O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 2011 Jun;43(6):585-589.

(319) Lambert CG, Black LJ. Learning from our GWAS mistakes: from experimental design to scientific method. Biostatistics 2012 Apr;13(2):195-203.

(320) Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 2010 Oct;11(10):733-739.

(321) Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet 2007 Sep;3(9):1724-1735.

(322) Pluzhnikov A, Below JE, Konkashbaev A, Tikhomirov A, Kistner-Griffin E, Roe CA, et al. Spoiling the whole bunch: quality control aimed at preserving the integrity of high-throughput genotyping. Am J Hum Genet 2010 Jul 9;87(1):123-128.

(323) Taub MA, Corrada Bravo H, Irizarry RA. Overcoming bias and systematic errors in next generation sequencing data. Genome Med 2010 Dec 10;2(12):87.

A.1. Marker information and nominal *p*-values for AD and ARPs for DA SNPs

GENE & CHR	SNP	BP	LOCATION	CALL RATE	MAF	AD (N=1028)	ONSET (N=436)	ISENS (N=428)	TOLMX (N=427)	MAX24 (N=436)	WDSFS (N=436)
DRD1	rs686	174801306	3' UTR	99.5	0.401	0.41	0.45	0.68	0.014	0.015	0.46
(Chr 5)	rs155417	174801446	exon 2	100.0	0.016	0.53	0.36	0.57	0.79	0.84	0.82
	rs5326	174802802	exon 2	98.8	0.128	0.80	0.83	0.72	0.20	0.18	0.35
	rs10078866	174804926	5' near gene	99.0	0.013	0.54	0.80	0.21	0.72	0.31	0.47
DRD2	rs2242592	112784640	3' near gene	98.1	0.275	0.10	0.24	0.83	0.52	0.03	0.33
(Chr 11)	rs6279	112786283	exon 8	99.2	0.279	0.13	0.27	0.78	0.59	0.02	0.33
	rs2587548	112797422	intron 1	99.4	0.389	0.30	0.24	0.15	0.98	0.10	0.64
	rs1076563	112801119	intron 1	99.7	0.389	0.35	0.19	0.17	0.97	0.11	0.58
	rs1079596	112801829	intron 1	99.5	0.162	0.36	0.82	0.37	0.97	0.83	0.39
	rs1125394	112802395	intron 1	99.8	0.160	0.45	0.79	0.36	0.95	0.76	0.35
	rs2471857	112803549	intron 1	100.0	0.162	0.29	0.84	0.32	0.94	0.98	0.41
	rs4648318	112818599	intron 1	99.9	0.239	0.34	0.15	0.73	0.62	0.16	0.69
	rs4274224	112824662	intron 1	99.3	0.482	0.11	0.47	0.33	0.57	0.84	0.89
	rs4581480	112829684	intron 1	99.7	0.109	0.13	0.69	0.65	0.26	0.02	0.98
	rs4648317	112836742	intron 1	99.9	0.149	0.79	0.80	0.92	0.25	0.46	0.39
	rs4630328	112839419	intron 1	93.1	0.391	0.09	0.56	0.06	0.11	0.26	0.49
	rs4350392	112840927	intron 1	98.9	0.161	0.31	0.65	0.87	0.22	0.32	0.39
	rs1799978	112851561	5' near gene	99.8	0.041	0.33	0.79	0.22	0.64	0.70	0.79
	rs12364283	112852165	5' near gene	99.9	0.079	0.64	0.86	0.33	0.29	0.71	0.68
DRD3	**rs2654754**	**115338486**	**intron 5**	**99.8**	**0.025**	0.30	1.00	**0.0021**	0.07	0.18	0.09
(Chr 3)	rs2134655	115340891	intron 5	99.7	0.273	0.13	0.71	0.44	0.99	0.15	0.89
	rs9288993	115341863	intron 4	99.9	0.023	0.37	0.78	0.00	0.07	0.18	0.07
	rs963468	115345577	intron 4	99.1	0.388	0.60	0.10	0.17	0.39	0.77	0.65
	rs3773678	115352768	intron 3	99.8	0.121	0.57	0.29	0.32	0.55	0.11	0.23
	rs2630349	115356062	intron 3	98.6	0.057	0.96	0.15	0.83	0.84	0.70	0.81
	rs167771	115358965	intron 3	99.8	0.156	0.18	0.14	0.55	0.28	0.56	0.09
	rs167770	115362252	intron 2	99.5	0.287	0.01	0.59	0.32	0.27	0.06	0.57
	rs324029	115364313	intron 2	99.4	0.288	0.01	0.75	0.22	0.24	0.04	0.61
	rs10934256	115368342	intron 2	98.9	0.211	0.15	1.00	0.47	0.13	0.10	0.45
	rs1486009	115371222	intron 2	99.9	0.050	0.09	0.43	0.52	0.08	0.85	0.33
	rs6280	115373505	exon 2	99.8	0.340	0.03	0.40	0.25	0.34	0.08	0.65
	rs9825563	115382910	5' near gene	99.0	0.317	0.35	0.57	0.22	0.14	0.07	0.68
DRD4	**rs12280580**	**616220**	**5' near gene**	**92.7**	**0.354**	**0.011**	0.66	0.59	0.55	0.17	0.76
(Chr 11)	rs3758653	626399	5' near gene	100.0	0.168	0.77	0.89	0.49	0.54	0.54	0.71
	rs916457	627014	5' near gene	100.0	0.044	0.22	0.74	0.56	0.81	0.49	0.78
	rs11246226	631191	3' near gene	99.2	0.476	0.25	0.30	0.58	0.32	0.98	0.95
DRD5	**rs7655090**	**9374973**	**3' near gene**	**98.1**	**0.049**	0.60	0.33	0.63	0.16	0.15	**0.0017**
(Chr 4)	rs11731100	9376278	3' near gene	96.7	0.376	0.97	0.71	1.00	0.14	0.27	0.22
	rs10939507	9377921	3' near gene	97.9	0.330	0.85	0.43	0.75	0.048	0.23	0.81
	rs2867383	9397033	3' near gene	99.7	0.332	0.89	0.62	0.75	0.06	0.24	0.72
	rs12500086	9418957	3' near gene	97.4	0.193	0.50	0.55	0.11	0.05	0.23	0.53
	rs10005226	9420903	3' near gene	98.3	0.193	0.73	0.50	0.34	0.73	0.41	0.85
	rs1401438	9423554	3' near gene	96.0	0.166	0.77	0.71	0.05	0.012	0.19	0.26
SLC18A2	rs363332	118992657	intron 2	100.0	0.259	0.72	0.09	0.72	0.59	0.31	0.12
(Chr 10)	rs363334	118994985	intron 3	99.4	0.256	0.89	0.09	0.74	0.61	0.32	0.13
	rs363338	118999379	intron 3	99.9	0.334	0.64	0.09	0.64	0.97	0.63	0.45
	rs3753127	118999902	intron 3	99.7	0.022	0.03	0.48	0.31	0.89	0.09	0.95
	rs2283139	119008847	intron 10	100.0	0.012	0.87	0.70	0.63	0.14	0.32	0.94
	rs4752045	119009680	intron 10	98.7	0.389	0.97	0.10	0.96	0.18	0.29	0.51
	rs363256	119014215	intron 10	99.8	0.093	0.65	0.25	0.35	0.05	0.66	0.37
	rs363271	119017401	intron 13	99.7	0.119	0.29	0.019	0.24	0.24	0.85	0.23
	rs363230	119019505	intron 13	99.8	0.406	0.38	0.33	0.75	0.05	0.49	0.30
	rs2244249	119022265	intron 15	99.6	0.141	0.51	0.00503	0.55	0.30	0.81	0.23
	rs363276	119023799	intron 15	99.8	0.142	0.56	0.01	0.37	0.33	0.93	0.25
SLC6A3	rs27072	1447522	intron 15	90.6	0.172	0.85	0.08	0.99	0.97	0.25	0.12
(Chr 5)	rs6869645	1457548	intron 12	100.0	0.044	0.55	0.98	0.89	0.44	0.04	0.77
	rs6347	1464412	exon 9	98.6	0.281	0.14	0.96	0.22	0.19	0.03	0.72
	rs27048	**1465645**	**intron 8**	**97.1**	**0.451**	**0.042**	0.68	0.24	0.61	0.79	0.24
	rs37022	1468629	intron 7	99.9	0.201	0.94	0.55	0.56	0.14	0.20	0.68
	rs11564758	1473588	intron 6	95.6	0.372	0.63	0.63	0.57	0.71	0.34	0.13
	rs464049	1476905	intron 4	99.8	0.460	0.68	0.71	0.50	0.68	0.05	0.64
	rs10052016	**1481111**	**intron 4**	**93.9**	0.412	**0.00055**	0.69	0.19	0.13	0.73	0.30
	rs460000	1485825	intron 3	99.4	0.230	0.74	0.75	0.42	0.83	0.01	0.64
	rs403636	1491354	intron 3	99.9	0.166	0.51	0.90	0.04	0.26	0.07	0.51
	rs6350	1496199	exon 2	99.9	0.076	0.90	0.66	0.76	0.34	0.16	0.03
	rs3756450	1501148	5' near gene	99.9	0.136	0.83	0.73	0.20	0.36	0.37	0.92

DDC (Chr 7)	rs4947510	50299629	3' near gene	99.2	0.486	0.62	0.46	0.13	0.64	0.51	0.58
	rs11575542	50305196	exon 14	99.7	0.015	0.31	0.57	0.37	0.76	0.68	0.64
	rs4947535	50305890	intron 13	99.5	0.360	0.30	0.86	0.07	0.92	0.34	0.34
	rs11761683	50317088	intron 11	99.8	0.325	0.62	0.94	0.30	0.43	0.28	0.93
	rs745043	50318164	intron 11	99.8	0.025	0.85	0.60	0.20	0.94	0.21	0.56
	rs732215	50318272	intron 11	97.6	0.285	0.64	0.83	0.24	0.56	0.68	0.66
	rs4490786	50318523	intron 11	93.0	0.358	0.23	0.69	0.97	0.44	0.64	0.13
	rs2122822	50326361	intron 9	95.5	0.199	0.29	0.82	0.22	0.86	0.55	0.82
	rs1037351	50339613	intron 8	99.7	0.433	0.50	0.65	0.71	0.36	0.92	0.39
	rs880028	50344345	intron 7	99.2	0.238	0.32	0.80	0.84	0.48	0.66	0.17
	rs11238178	50365599	intron 6	100.0	0.437	0.13	0.70	0.08	0.89	0.04	0.68
	rs11575343	50370556	intron 5	95.2	0.342	0.48	0.86	0.46	0.66	0.90	0.84
	rs10244632	50372912	intron 4	99.5	0.240	0.16	0.77	0.20	0.83	0.07	0.76
	rs1466163	50381415	intron 3	98.2	0.269	0.59	0.58	0.67	0.53	0.11	0.67
	rs7786398	50387115	intron 1	100.0	0.022	0.40	0.50	0.55	0.85	0.94	0.69
	rs10499695	50392813	intron 1	98.8	0.293	0.12	0.34	0.89	0.96	0.30	0.66
	rs2329341	50394484	intron 1	99.3	0.109	0.45	0.55	0.37	0.98	0.92	0.20
	rs10499696	50395797	intron 1	94.6	0.482	0.33	0.95	0.69	0.66	0.14	0.49
	rs921451	50397494	intron 1	94.2	0.478	0.51	0.61	0.50	0.73	0.87	0.26
	rs6969081	50398714	intron 1	99.7	0.358	0.18	0.75	0.24	0.77	0.16	0.52
	rs3829897	50403973	intron 1	99.7	0.116	0.06	0.35	0.83	0.42	0.54	0.26
	rs7804365	50411357	5' near gene	99.7	0.353	0.04	0.27	0.94	0.66	0.21	0.11
TH (Chr 11)	rs2070762	2142911	intron 13	91.9	0.403	0.22	0.93	0.50	0.67	0.88	0.91
	rs11564717	**2143465**	**intron 12**	**99.9**	**0.408**	0.31	0.70	0.20	0.91	0.05	**0.0094**
	rs6356	2147527	exon 3	98.9	0.479	0.14	0.53	0.32	0.75	0.98	0.56
COMT (Chr 22)	rs737866	18304663	intron 1	99.7	0.286	0.29	0.73	0.33	0.04	0.37	0.68
	rs933271	18305961	intron 1	99.6	0.282	0.02	0.96	0.06	0.04	0.65	0.44
	rs5993883	18312192	intron 1	99.8	0.499	0.06	0.93	0.40	0.24	0.40	0.52
	rs740603	18319731	intron 1	99.6	0.487	0.17	0.92	0.42	0.36	0.25	0.63
	rs2239393	18324982	intron 3	99.2	0.407	0.61	1.00	0.88	0.99	0.28	0.77
	rs4680	18325825	exon 4	99.4	0.498	0.66	0.38	0.84	0.98	0.51	0.26
	rs4646316	18326686	intron 5	98.6	0.260	0.45	0.46	0.98	0.58	0.85	0.11
	rs174696	18327730	intron 5	100.0	0.219	0.86	0.14	0.80	0.23	0.02	0.48
	rs174697	18328386	intron 5	99.1	0.061	0.42	0.90	0.88	0.66	0.29	0.61
	rs9332377	18330246	intron 5	99.8	0.150	0.29	0.72	0.87	0.97	0.38	0.22

Notes: AD, alcohol dependence; ONSET, age-at-onset of AD; ISENS, initial sensitivity; TOLMX, tolerance/maximum drinking; MAX24, maximum drinks in 24 hours; WDSFS, withdrawal severity factor score. Nominal *p*-values significant after permutation are in bold.

A.2. Marker information and nominal *p*-values for disinhibitory symptom counts and factor score for DA SNPs

GENE & CHR	SNP	BP	LOCATION	CALL RATE	MAF	ADsx (N=436)	ADHDsx (N=426)	ASPDsx (N=436)	CDsx (N=436)	DDsx (N=432)	NS (N=429)	Factor Score (N=426)
DRD1	rs686	174801306	3' UTR	99.5	0.401	0.24	0.36	0.83	0.77	0.58	0.74	0.37
(Chr 5)	rs155417	174801446	exon 2	100.0	0.016	0.20	0.03	0.43	0.59	0.86	0.54	0.26
	rs5326	174802802	exon 2	98.8	0.128	0.31	0.20	0.19	0.07	0.07	0.95	0.05
	rs10078866	174804926	5' near gene	99.0	0.013	0.59	0.38	0.021	0.023	0.75	0.46	0.02
DRD2	rs2242592	112784640	3' near gene	98.1	0.275	0.60	0.72	0.60	0.73	0.22	0.42	0.43
(Chr 11)	rs6279	112786283	exon 8	99.2	0.279	0.74	0.58	0.53	0.64	0.26	0.43	0.35
	rs2587548	112797422	intron 1	99.4	0.389	0.80	0.34	0.85	0.71	0.52	0.047	0.35
	rs1076563	112801119	intron 1	99.7	0.389	0.83	0.45	0.91	0.76	0.49	0.05	0.40
	rs1079596	112801829	intron 1	99.5	0.162	0.83	0.28	0.43	0.59	0.32	0.29	0.84
	rs1125394	112802395	intron 1	99.8	0.160	0.68	0.31	0.46	0.59	0.30	0.31	0.81
	rs2471857	112803549	intron 1	100.0	0.162	0.77	0.24	0.50	0.71	0.37	0.26	0.94
	rs4648318	112818599	intron 1	99.9	0.239	0.69	0.40	0.82	0.87	0.30	0.08	0.62
	rs4274224	112824662	intron 1	99.3	0.482	0.87	0.93	0.19	0.72	0.43	0.028	0.71
	rs4581480	112829684	intron 1	99.7	0.109	0.89	0.42	0.66	0.46	0.39	0.07	0.21
	rs4648317	112836742	intron 1	99.9	0.149	0.58	0.39	0.82	0.17	0.21	0.82	0.09
	rs4630328	112839419	intron 1	93.1	0.391	0.91	0.50	0.12	0.84	0.47	0.00049	1.00
	rs4350392	112840927	intron 1	98.9	0.161	0.44	0.24	0.81	0.14	0.21	0.73	0.07
	rs1799978	112851561	5' near gene	99.8	0.041	0.85	0.29	0.73	0.62	0.51	0.54	0.47
	rs12364283	112852165	5' near gene	99.9	0.079	0.17	0.64	0.09	0.78	0.91	0.42	0.84
DRD3	rs2654754	115338486	intron 5	99.8	0.025	0.05	0.55	0.25	0.23	0.84	0.84	0.29
(Chr 3)	rs2134655	115340891	intron 5	99.7	0.273	0.84	0.81	0.63	0.85	0.57	0.58	0.56
	rs9288993	115341863	intron 4	99.9	0.023	0.07	0.55	0.18	0.25	0.91	0.79	0.28
	rs963468	115345577	intron 4	99.1	0.388	0.70	0.37	0.29	0.88	0.51	0.74	0.66
	rs3773678	115352768	intron 3	99.8	0.121	0.97	0.65	0.79	0.85	0.06	0.73	0.78
	rs2630349	115356062	intron 3	98.6	0.057	0.97	0.73	0.59	0.50	0.14	0.77	0.50
	rs167771	115358965	intron 3	99.8	0.156	0.99	0.12	0.70	0.87	0.05	0.83	1.00
	rs167770	115362252	intron 2	99.5	0.287	0.43	0.78	0.32	0.92	0.67	0.49	0.95
	rs324029	115364313	intron 2	99.4	0.288	0.39	0.58	0.25	0.68	0.65	0.54	0.77
	rs10934256	115368342	intron 2	98.9	0.211	0.11	0.46	0.52	0.90	0.89	0.79	0.80
	rs1486009	115371222	intron 2	99.9	0.050	0.85	0.06	0.60	0.97	0.35	0.38	0.89
	rs6280	115373505	exon 2	99.8	0.340	0.43	0.72	0.87	0.52	0.38	0.38	0.60
	rs9825563	115382910	5' near gene	99.0	0.317	0.44	0.68	0.63	0.80	0.51	0.99	0.98
DRD4	rs12280580	616220	5' near gene	92.7	0.354	0.81	0.93	0.52	0.92	0.56	0.47	0.86
(Chr 11)	rs3758653	626399	5' near gene	100.0	0.168	0.90	0.92	0.87	0.79	0.46	0.80	0.87
	rs916457	627014	5' near gene	100.0	0.044	0.32	0.21	0.03	0.10	0.90	0.68	0.22
	rs11246226	631191	3' near gene	99.2	0.476	0.65	0.45	0.92	0.81	0.42	0.45	0.89
DRD5	**rs7655090**	**9374973**	**3' near gene**	**98.1**	**0.049**	0.15	0.08	0.17	0.029	0.07	0.38	**0.0094**
(Chr 4)	rs11731100	9376278	3' near gene	96.7	0.376	0.10	0.90	0.08	0.0063	0.48	0.78	0.09
	rs10939507	9377921	3' near gene	97.9	0.330	0.58	0.34	0.21	0.06	0.71	0.79	0.55
	rs2867383	9397033	3' near gene	99.7	0.332	0.50	0.53	0.05	0.016	0.79	0.82	0.28
	rs12500086	9418957	3' near gene	97.4	0.193	0.48	0.77	0.15	0.21	0.88	0.50	0.60
	rs10005226	9420903	3' near gene	98.3	0.193	0.18	0.04	0.94	0.59	0.67	0.14	0.40
	rs1401438	9423554	3' near gene	96.0	0.166	0.16	0.84	0.18	0.17	0.81	0.75	0.79
SLC18A2	rs363332	118992657	intron 2	100.0	0.259	0.016	0.13	0.21	0.71	0.67	0.19	0.18
(Chr 10)	rs363334	118994985	intron 3	99.4	0.256	0.02	0.13	0.22	0.73	0.67	0.18	0.19
	rs363338	118999379	intron 3	99.9	0.334	0.13	0.30	0.24	0.47	0.36	0.23	0.32
	rs3753127	118999902	intron 3	99.7	0.022	0.92	0.94	0.62	0.99	0.27	0.57	0.84
	rs2283139	119008847	intron 10	100.0	0.012	0.20	0.79	0.04	0.24	0.51	0.29	0.19
	rs4752045	119009680	intron 10	98.7	0.389	0.33	0.25	0.72	0.48	0.80	0.35	0.86
	rs363256	119014215	intron 10	99.8	0.093	0.69	0.76	0.08	0.27	0.40	0.07	0.15
	rs363271	119017401	intron 13	99.7	0.119	0.98	0.62	0.09	0.20	0.48	0.35	0.32
	rs363230	119019505	intron 13	99.8	0.406	0.77	0.69	0.55	0.84	0.29	0.41	0.51
	rs2244249	119022265	intron 15	99.6	0.141	0.60	0.68	0.08	0.27	0.49	0.34	0.54
	rs363276	119023799	intron 15	99.8	0.142	0.64	0.65	0.09	0.31	0.55	0.36	0.55
SLC6A3	rs27072	1447522	intron 15	90.6	0.172	0.84	0.29	0.89	0.90	0.45	1.00	0.96
(Chr 5)	rs6869645	1457548	intron 12	100.0	0.044	0.19	0.86	0.97	0.56	0.84	0.76	0.80

A.2. Marker information and nominal *p*-values for disinhibitory symptom counts and factor score for DA SNPs

GENE & CHR	SNP	BP	LOCATION	CALL RATE	MAF	ADsx (N=436)	ADHDsx (N=426)	ASPDsx (N=436)	CDsx (N=436)	DDsx (N=432)	NS (N=429)	Factor Score (N=426)
	rs6347	1464412	exon 9	98.6	0.281	0.57	0.70	0.18	0.21	0.61	0.67	0.31
	rs27048	1465645	intron 8	97.1	0.451	0.74	0.90	0.35	0.45	0.64	0.43	0.94
	rs37022	1468629	intron 7	99.9	0.201	0.32	0.38	0.60	0.89	0.76	0.96	0.72
	rs11564758	1473588	intron 6	95.6	0.372	0.92	0.38	0.15	0.98	0.56	0.93	1.00
	rs464049	1476905	intron 4	99.8	0.460	0.93	0.96	0.75	0.71	0.73	0.70	0.88
	rs10052016	1481111	intron 4	93.9	0.412	0.81	0.44	0.37	0.06	0.57	0.25	0.60
	rs460000	1485825	intron 3	99.4	0.230	0.39	0.17	0.56	0.63	0.83	0.70	0.96
	rs403636	1491354	intron 3	99.9	0.166	0.83	0.37	0.46	0.85	0.86	0.85	0.99
	rs6350	**1496199**	**exon 2**	**99.9**	**0.076**	**0.0021**	0.94	0.52	0.80	0.89	0.31	0.89
	rs3756450	1501148	5' near gene	99.9	0.136	0.51	0.08	0.69	0.85	0.03	0.58	0.82
DDC	rs4947510	50299629	3' near gene	99.2	0.486	0.99	0.12	0.44	0.83	0.16	0.27	0.52
(Chr 7)	**rs11575542**	**50305196**	**exon 14**	**99.7**	**0.015**	0.22	0.48	0.54	0.63	**0.0028**	0.86	0.52
	rs4947535	50305890	intron 13	99.5	0.360	0.55	0.44	0.60	0.54	0.65	0.64	0.56
	rs11761683	50317088	intron 11	99.8	0.325	0.41	0.86	0.28	0.50	0.24	0.69	0.90
	rs745043	50318164	intron 11	99.8	0.025	1.00	0.79	0.97	0.58	0.88	0.32	1.00
	rs732215	50318272	intron 11	97.6	0.285	0.97	0.28	0.98	0.55	0.40	0.33	0.55
	rs4490786	50318523	intron 11	93.0	0.358	0.40	0.16	0.21	0.87	0.17	0.08	0.17
	rs2122822	50326361	intron 9	95.5	0.199	0.30	0.19	0.66	0.76	0.95	0.24	0.55
	rs1037351	50339613	intron 8	99.7	0.433	0.46	0.72	0.73	0.61	0.79	0.78	0.70
	rs880028	50344345	intron 7	99.2	0.238	0.37	0.19	0.23	0.88	0.23	0.10	0.19
	rs11238178	50365599	intron 6	100.0	0.437	0.50	0.43	0.28	0.31	0.93	0.87	0.17
	rs11575343	50370556	intron 5	95.2	0.342	0.34	0.97	0.39	0.90	0.01	0.63	0.29
	rs10244632	50372912	intron 4	99.5	0.240	0.38	0.50	0.26	0.45	0.44	0.86	0.12
	rs1466163	50381415	intron 3	98.2	0.269	0.32	0.84	0.46	0.80	0.73	0.66	0.81
	rs7786398	50387115	intron 1	100.0	0.022	0.23	0.44	0.06	0.86	0.95	0.21	0.08
	rs10499695	50392813	intron 1	98.8	0.293	0.49	0.88	0.21	0.73	0.27	0.28	0.97
	rs2329341	50394484	intron 1	99.3	0.109	0.41	0.45	0.61	0.82	0.43	0.85	0.60
	rs10499696	50395797	intron 1	94.6	0.482	0.44	0.76	0.76	0.80	0.82	0.43	0.98
	rs921451	50397494	intron 1	94.2	0.478	0.53	0.31	0.53	0.75	0.48	0.83	0.46
	rs6969081	50398714	intron 1	99.7	0.358	0.79	0.82	0.51	0.41	0.55	0.83	0.98
	rs3829897	50403973	intron 1	99.7	0.116	0.70	0.38	0.08	0.27	0.32	0.93	0.12
	rs7804365	50411357	5' near gene	99.7	0.353	0.95	0.83	0.22	0.35	0.41	0.41	0.23
TH	rs2070762	2142911	intron 13	91.9	0.403	0.68	0.43	0.50	0.98	0.51	0.56	0.96
(Chr 11)	rs11564717	2143465	intron 12	99.9	0.408	0.16	0.014[a]	0.46	0.27	0.19	0.17	0.13
	rs6356	2147527	exon 3	98.9	0.479	0.80	0.98	0.85	0.83	0.50	0.52	0.63
COMT	rs737866	18304663	intron 1	99.7	0.286	0.97	0.03	0.52	0.90	0.08	0.54	0.39
(Chr 22)	rs933271	18305961	intron 1	99.6	0.282	0.78	0.13	0.44	0.64	0.53	0.69	0.91
	rs5993883	18312192	intron 1	99.8	0.499	0.85	0.57	0.53	0.74	0.02	0.62	0.32
	rs740603	18319731	intron 1	99.6	0.487	0.95	0.92	0.81	0.79	0.05	0.89	0.60
	rs2239393	18324982	intron 3	99.2	0.407	0.14	0.42	0.99	0.63	0.52	0.33	0.97
	rs4680	18325825	exon 4	99.4	0.498	0.54	0.79	0.49	0.12	0.55	0.26	0.35
	rs4646316	18326686	intron 5	98.6	0.260	0.55	0.44	0.44	0.59	0.76	0.14	0.69
	rs174696	18327730	intron 5	100.0	0.219	0.49	0.72	0.11	0.18	0.74	0.75	0.22
	rs174697	18328386	intron 5	99.1	0.061	0.99	0.13	0.71	0.88	0.46	0.58	0.48
	rs9332377	18330246	intron 5	99.8	0.150	0.11	0.83	0.79	0.51	0.88	0.13	0.88

Notes: ADsx, alcohol dependence symptoms; APSDsx, antisocial personality disorder symptoms; ADHDsx, attention deficit/hyperactivity symptoms; CDsx, conduct disorder symptoms; DDsx, drug dependence symptoms; NS, novelty seeking score. Nominal *p*-values significant after permutation are in bold.

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs2154358	1	20143206	0.92	1.09E-05	0.3	0.92	1.0	intergenic
rs111575277	1	47789513	0.73	1.10E-05	0.3	0.44	0.9	intergenic
rs2182241	1	62805139	1.00	4.11E-06	0.2	0.68	1.0	intergenic
rs76132455	1	64839483	0.56	1.82E-05	0.4	0.63	1.0	intergenic
rs59409703	1	65535046	0.72	3.11E-06	0.2	0.49	0.9	intergenic
rs3762312	1	65535986	0.72	2.42E-06	0.2	0.49	0.9	intergenic
rs114102876	1	65560909	0.72	9.60E-06	0.3	0.28	0.7	intergenic
rs61798772	1	89641171	0.88	1.52E-05	0.3	0.77	1.0	*GBP7*
rs139254361	1	89732527	0.91	7.32E-07	0.2	0.64	1.0	*GBP5*
rs58778527	1	100407807	0.99	3.14E-05	0.5	0.93	1.0	intergenic
rs1541044	1	100411287	0.99	3.90E-05	0.5	0.93	1.0	intergenic
rs1530683	1	104990673	0.95	2.06E-05	0.4	0.15	0.6	intergenic
rs7417007	1	105002929	0.99	2.20E-05	0.4	0.16	0.6	intergenic
rs9662391	1	105014143	0.99	6.11E-06	0.2	0.22	0.7	intergenic
rs72987611	1	105020384	0.99	5.75E-06	0.2	0.14	0.6	intergenic
rs74105009	1	105029400	0.98	3.66E-06	0.2	0.11	0.6	intergenic
rs9662087	1	105047970	0.99	3.52E-06	0.2	0.21	0.7	intergenic
rs6583098	1	105050992	0.99	4.31E-06	0.2	0.22	0.7	intergenic
rs75382510	1	105054941	0.98	2.13E-06	0.2	0.21	0.7	intergenic
rs75182375	1	105055795	0.99	4.39E-06	0.2	0.23	0.7	intergenic
rs74634766	1	105056575	0.99	4.39E-06	0.2	0.24	0.7	intergenic
rs2218672	1	105060691	0.98	4.45E-06	0.2	0.25	0.7	intergenic
rs76565683	1	105061521	0.98	4.48E-06	0.2	0.25	0.7	intergenic
rs1448393	1	105115585	0.99	4.15E-05	0.5	0.28	0.7	intergenic
rs11589124	1	105116823	0.99	4.15E-05	0.5	0.28	0.7	intergenic
rs74557654	1	105119188	0.96	4.06E-05	0.5	0.30	0.7	intergenic
rs11589477	1	105142710	0.93	3.97E-05	0.5	0.12	0.6	intergenic
rs6662838	1	156403701	0.63	2.31E-05	0.4	0.78	1.0	intergenic
rs11590002	1	175032518	1.00	2.17E-06	0.2	0.30	0.7	intergenic
rs11582546	1	175049009	0.91	2.58E-06	0.2	0.03	0.6	*TNN*
rs61827435	1	175066967	0.90	2.44E-06	0.2	0.02	0.6	*TNN*
rs59698324	1	201965855	0.99	3.51E-05	0.5	0.90	1.0	*RNPEP*
rs61821542	1	201967669	1.00	2.81E-05	0.4	0.98	1.0	*RNPEP*
rs13375435	1	201971862	1.00	3.02E-05	0.5	0.99	1.0	*RNPEP*
rs4630172	1	201972889	1.00	3.03E-05	0.5	0.99	1.0	*RNPEP*
rs3820439	1	201973565	1.00	3.03E-05	0.5	0.97	1.0	*RNPEP*
rs7516412	1	201976201	0.99	3.57E-05	0.5	0.95	1.0	LOC1002894
rs36073842	1	204694428	0.51	4.35E-05	0.5	0.15	0.6	intergenic
rs114901102	1	210901331	0.89	9.06E-06	0.3	0.13	0.6	*KCNH1*
rs16845712	1	226536337	0.75	8.25E-07	0.2	0.46	0.9	intergenic
rs6701940	1	226649170	0.72	1.66E-05	0.4	0.05	0.6	intergenic
rs185448	1	238471801	0.99	2.33E-05	0.4	0.82	1.0	intergenic
rs259578	1	238471842	0.99	2.36E-05	0.4	0.82	1.0	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs259577	1	238471915	0.99	2.42E-05	0.4	0.87	1.0	intergenic
rs259602	1	238481474	1.00	2.73E-05	0.4	0.89	1.0	intergenic
rs259601	1	238483478	1.00	2.72E-05	0.4	0.88	1.0	intergenic
rs259580	1	238492061	1.00	1.67E-05	0.4	0.82	1.0	intergenic
rs7539749	1	238492486	0.99	1.72E-05	0.4	0.70	1.0	intergenic
rs1414538	1	238498167	0.99	9.01E-06	0.3	0.87	1.0	intergenic
rs1027189	1	238500210	1.00	1.57E-05	0.4	0.83	1.0	intergenic
rs1964602	1	238501171	0.97	2.90E-05	0.5	0.74	1.0	intergenic
rs1361655	1	238504548	1.00	1.22E-05	0.3	0.91	1.0	intergenic
rs7538546	1	238512343	1.00	1.22E-05	0.3	0.92	1.0	intergenic
rs7550834	1	238516057	1.00	1.21E-05	0.3	0.86	1.0	intergenic
rs4659522	1	238518836	0.99	1.08E-05	0.3	0.94	1.0	intergenic
rs10737824	1	238519408	0.99	4.00E-05	0.5	1.00	1.0	intergenic
rs6702470	1	238520218	0.99	3.91E-05	0.5	1.00	1.0	intergenic
rs7550792	1	238522218	0.99	8.15E-06	0.3	1.00	1.0	intergenic
rs10158750	1	238525963	0.95	4.18E-06	0.2	0.96	1.0	intergenic
rs12129735	1	238528907	0.85	2.59E-05	0.4	0.79	1.0	intergenic
rs72768910	2	15940724	0.94	2.95E-05	0.5	0.09	0.6	intergenic
rs77224371	2	16948876	0.84	1.76E-05	0.4	0.39	0.8	intergenic
rs10203610	2	27572265	0.74	1.71E-05	0.4	0.05	0.6	*GTF3C2*
rs74467769	2	35410184	0.66	3.00E-05	0.5	NA	NA	intergenic
rs10496077	2	57943044	0.96	3.39E-05	0.5	0.87	1.0	intergenic
rs55950043	2	57943869	0.82	1.30E-05	0.3	0.73	1.0	intergenic
rs72804544	2	57970078	0.83	2.66E-05	0.4	0.82	1.0	intergenic
rs72804548	2	57974646	0.88	1.12E-05	0.3	0.67	1.0	intergenic
rs72804553	2	57982530	0.82	4.04E-06	0.2	0.78	1.0	intergenic
rs67973994	2	57987808	0.99	3.97E-05	0.5	0.59	1.0	intergenic
rs72804580	2	58048234	0.84	1.61E-05	0.4	0.83	1.0	intergenic
rs72840080	2	85596043	0.79	2.59E-05	0.4	0.38	0.8	*ELMOD3*
rs112480835	2	104070773	0.56	3.09E-05	0.5	0.39	0.8	intergenic
rs72833664	2	114506030	0.97	3.07E-07	0.1	NA	NA	*SLC35F5*
rs113069166	2	115918943	0.90	9.60E-07	0.2	0.73	1.0	*DPP10*
rs112287044	2	116086394	1.00	4.40E-05	0.5	0.73	1.0	*DPP10*
rs17823041	2	116101097	1.00	4.12E-05	0.5	0.44	0.9	*DPP10*
rs66827093	2	116160527	0.89	5.36E-06	0.2	0.98	1.0	*DPP10*
rs150113144	2	122393805	0.83	2.52E-05	0.4	0.46	0.9	*CLASP1*
rs17753709	2	145986219	0.50	1.59E-05	0.4	0.62	1.0	intergenic
rs142227413	2	154351630	0.85	2.17E-05	0.4	0.50	0.9	intergenic
rs72920194	2	199362905	0.72	6.62E-06	0.2	0.74	1.0	intergenic
rs191005325	2	199392149	0.99	1.52E-06	0.2	NA	NA	intergenic
rs137943609	2	199462960	0.97	1.35E-06	0.2	NA	NA	intergenic
rs17229679	2	199560757	0.64	9.26E-06	0.3	0.91	1.0	intergenic
rs116203419	2	200649995	0.77	1.39E-05	0.3	0.19	0.6	*FONG*

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs193137425	2	206245058	0.57	3.21E-05	0.5	0.65	1.0	*PARD3B*
rs79260980	2	215012873	0.72	2.93E-05	0.5	0.65	1.0	*SPAG16*
rs77454352	2	222078482	0.91	2.71E-05	0.4	0.14	0.6	intergenic
rs28439910	2	238209899	0.91	1.68E-05	0.4	0.09	0.6	intergenic
rs28432742	2	238210290	0.97	1.37E-05	0.3	0.10	0.6	intergenic
rs28620087	2	238210323	0.97	1.37E-05	0.3	0.10	0.6	intergenic
rs28666656	2	238210395	0.98	1.37E-05	0.3	0.10	0.6	intergenic
rs28497536	2	238210416	0.98	1.38E-05	0.3	0.10	0.6	intergenic
rs11897292	2	238210640	1.00	1.30E-05	0.3	0.11	0.6	intergenic
rs11897373	2	238210881	1.00	1.29E-05	0.3	0.09	0.6	intergenic
rs13404258	2	238214571	1.00	1.38E-05	0.3	0.10	0.6	intergenic
rs6729500	2	238215479	1.00	1.74E-05	0.4	0.07	0.6	intergenic
rs7598750	2	238217274	1.00	1.54E-05	0.4	0.07	0.6	intergenic
rs10184009	2	238219581	1.00	1.88E-05	0.4	0.06	0.6	intergenic
rs61035261	2	238221360	1.00	2.07E-05	0.4	0.06	0.6	intergenic
rs10929224	2	238225610	1.00	2.29E-05	0.4	0.04	0.6	intergenic
rs10204947	2	238226513	0.99	2.26E-05	0.4	0.05	0.6	intergenic
rs924209	2	238227536	0.99	2.10E-05	0.4	0.05	0.6	intergenic
rs56310758	2	238240863	1.00	1.25E-08	0.1	0.04	0.6	*COL6A3*
rs75561681	2	238241881	0.97	2.77E-08	0.1	0.01	0.6	*COL6A3*
rs2270671	2	238243285	1.00	1.18E-08	0.1	0.06	0.6	*COL6A3*
rs10929226	2	238244559	0.99	1.17E-08	0.1	0.06	0.6	*COL6A3*
rs112523013	2	238247257	0.99	1.05E-08	0.1	0.07	0.6	*COL6A3*
rs2646258	2	238253149	0.99	6.39E-09	0.1	0.05	0.6	*COL6A3*
rs2256485	2	238253930	0.99	6.18E-09	0.1	0.03	0.6	*COL6A3*
rs2270656	2	238255120	0.99	9.12E-07	0.2	0.13	0.6	*COL6A3*
rs2645767	2	238256171	1.00	1.60E-06	0.2	0.24	0.7	*COL6A3*
rs2646265	2	238257013	1.00	1.46E-08	0.1	0.07	0.6	*COL6A3*
rs3790990	2	238257127	1.00	2.45E-05	0.4	0.30	0.7	*COL6A3*
rs2646264	2	238257213	1.00	1.19E-08	0.1	0.10	0.6	*COL6A3*
rs2646263	2	238257353	0.99	1.56E-06	0.2	0.19	0.6	*COL6A3*
rs2645766	2	238257896	0.99	2.38E-05	0.4	0.31	0.8	*COL6A3*
rs2646261	2	238259387	1.00	2.36E-08	0.1	0.08	0.6	*COL6A3*
rs3790999	2	238260701	1.00	1.83E-06	0.2	0.38	0.8	*COL6A3*
rs3828135	2	238260829	0.99	1.03E-05	0.3	0.69	1.0	*COL6A3*
rs11901326	2	238261509	0.99	5.00E-07	0.1	0.33	0.8	*COL6A3*
rs7605340	2	238261744	0.99	1.12E-05	0.3	0.42	0.9	*COL6A3*
rs7605341	2	238261746	0.99	9.53E-06	0.3	0.39	0.8	*COL6A3*
rs2646257	2	238261850	0.99	5.58E-07	0.1	0.33	0.8	*COL6A3*
rs2645764	2	238262254	1.00	7.09E-09	0.1	0.18	0.6	*COL6A3*
rs2645763	2	238263299	1.00	7.97E-09	0.1	0.17	0.6	*COL6A3*
rs10929228	2	238265684	0.96	1.62E-05	0.4	0.40	0.8	*COL6A3*
rs2645777	2	238266146	0.96	4.41E-08	0.1	0.13	0.6	*COL6A3*

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs2646254	2	238267717	0.88	3.98E-05	0.5	0.32	0.8	*COL6A3*
rs2645769	2	238269871	0.96	5.11E-06	0.2	0.23	0.7	*COL6A3*
rs7597795	2	238270726	0.98	1.36E-05	0.3	0.24	0.7	*COL6A3*
rs10202497	2	238270894	0.96	1.43E-06	0.2	0.16	0.6	*COL6A3*
rs7598394	2	238271229	1.00	1.73E-05	0.4	0.31	0.8	*COL6A3*
rs10167850	2	238271284	0.97	2.35E-06	0.2	0.17	0.6	*COL6A3*
rs11915389	3	6960199	0.99	2.47E-05	0.4	0.20	0.6	*GRM7*
rs115937108	3	6992572	0.99	2.30E-05	0.4	0.23	0.7	*GRM7*
rs73808494	3	7002079	0.99	2.98E-05	0.5	0.27	0.7	*GRM7*
rs149456087	3	55912341	0.73	1.66E-05	0.4	0.46	0.9	*ERC2*
rs75104355	3	67223691	1.00	3.20E-05	0.5	0.44	0.9	intergenic
rs9870011	3	67232976	0.99	2.98E-05	0.5	0.39	0.8	intergenic
rs12635253	3	67237603	1.00	2.98E-05	0.5	0.29	0.7	intergenic
rs9844493	3	67261017	1.00	2.23E-05	0.4	0.29	0.7	intergenic
rs13067571	3	67263927	0.99	1.35E-05	0.3	0.35	0.8	intergenic
rs4857009	3	67267448	1.00	2.39E-05	0.4	0.39	0.8	intergenic
rs138167890	3	79572545	0.71	1.30E-05	0.3	NA	NA	*ROBO1*
rs111317410	3	79596496	0.71	1.67E-05	0.4	NA	NA	*ROBO1*
rs113065987	3	79616504	0.71	1.72E-05	0.4	NA	NA	*ROBO1*
rs112777194	3	79861124	0.71	1.82E-05	0.4	NA	NA	intergenic
rs113597039	3	79873473	0.70	1.51E-05	0.3	NA	NA	intergenic
rs139966448	3	79938342	0.70	1.45E-05	0.3	NA	NA	intergenic
rs185680021	3	80011168	0.56	1.67E-05	0.4	NA	NA	intergenic
rs188263218	3	80031296	0.63	3.96E-05	0.5	NA	NA	intergenic
rs149144768	3	80065406	0.70	1.14E-05	0.3	NA	NA	intergenic
rs190250423	3	96339264	0.48	1.75E-05	0.4	0.50	0.9	intergenic
rs74686325	3	107151825	0.86	2.76E-05	0.4	0.26	0.7	intergenic
rs8177279	3	133485366	0.83	4.45E-05	0.5	NA	NA	*TF*
rs147453475	3	141554677	0.83	2.08E-05	0.4	NA	NA	intergenic
rs73149932	3	147209249	0.83	8.39E-07	0.2	0.39	0.8	intergenic
rs73149934	3	147209715	0.82	6.23E-07	0.1	0.53	0.9	intergenic
rs183486479	3	148499451	0.52	5.44E-06	0.2	0.63	1.0	intergenic
rs78046234	3	149328575	0.69	3.74E-05	0.5	0.31	0.8	*WWTR1*
rs186341850	3	153682165	0.65	3.95E-05	0.5	0.58	1.0	intergenic
rs150268941	3	172575312	0.91	1.39E-08	0.1	0.20	0.6	intergenic
rs150327282	3	175334707	0.88	2.29E-06	0.2	NA	NA	*NAALADL2*
rs143156290	3	175367776	0.92	2.01E-06	0.2	NA	NA	*NAALADL2*
rs190711641	3	175376010	0.93	1.82E-06	0.2	NA	NA	*NAALADL2*
rs114114549	3	175392668	0.91	1.70E-06	0.2	0.77	1.0	*NAALADL2*
rs16826149	3	175402695	0.97	2.50E-05	0.4	0.96	1.0	*NAALADL2*
rs142645748	3	175451990	0.93	3.80E-07	0.1	NA	NA	*NAALADL2*
rs6776133	3	175458526	0.99	1.90E-06	0.2	0.96	1.0	*NAALADL2*
rs74201644	3	175459872	0.99	2.69E-06	0.2	1.00	1.0	*NAALADL2*

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs9857587	3	175463441	0.99	2.64E-06	0.2	0.98	1.0	*NAALADL2*
rs7609843	3	175485960	0.98	2.38E-06	0.2	0.98	1.0	*NAALADL2*
rs139426341	3	175503655	0.55	1.24E-05	0.3	0.39	0.8	*NAALADL2*
rs183579266	3	178477833	0.55	9.41E-07	0.2	0.29	0.7	*KCNMB2*
rs73185354	3	178673972	0.89	2.25E-05	0.4	0.87	1.0	intergenic
rs114261362	3	178676706	0.89	2.23E-05	0.4	0.87	1.0	intergenic
rs73185398	3	178722033	0.89	3.41E-05	0.5	0.80	1.0	intergenic
rs73185401	3	178723562	0.89	3.35E-05	0.5	0.80	1.0	intergenic
rs2268844	3	185920603	1.00	1.24E-05	0.3	0.48	0.9	*DGKG*
rs189732250	3	192855932	0.73	2.18E-05	0.4	NA	NA	intergenic
rs4689549	4	6776649	1.00	1.50E-05	0.3	0.64	1.0	intergenic
rs2324647	4	25040739	1.00	3.70E-05	0.5	0.82	1.0	intergenic
rs2324648	4	25041054	1.00	4.14E-05	0.5	0.82	1.0	intergenic
rs113600173	4	58649957	0.85	8.47E-06	0.3	0.05	0.6	intergenic
rs186460260	4	61621339	0.93	3.31E-05	0.5	NA	NA	intergenic
rs151098570	4	61660503	0.94	4.11E-05	0.5	NA	NA	intergenic
rs190727238	4	61701164	0.94	4.55E-06	0.2	NA	NA	intergenic
rs140667745	4	88451921	0.81	2.99E-05	0.5	0.63	1.0	intergenic
rs7657524	4	88742562	0.89	3.61E-05	0.5	0.55	0.9	intergenic
rs265070	4	96401647	1.00	4.53E-05	0.5	0.14	0.6	*UNC5C*
rs116306628	4	114327260	0.82	3.90E-05	0.5	0.97	1.0	intergenic
rs145313326	4	136556300	0.83	2.05E-05	0.4	NA	NA	intergenic
rs72687317	4	158224381	0.80	7.49E-06	0.3	0.63	1.0	*GRIA2*
rs72689009	4	158427649	0.90	5.49E-06	0.2	0.45	0.9	intergenic
rs114397078	4	176431598	0.70	3.92E-05	0.5	0.27	0.7	intergenic
rs116122755	4	183267177	0.81	9.60E-06	0.3	0.84	1.0	*ODZ3*
rs6823293	4	188314906	0.92	2.10E-05	0.4	0.03	0.6	LOC339975
rs78616487	4	188322087	0.96	1.84E-05	0.4	0.05	0.6	LOC339975
rs11943588	4	188322659	0.94	1.20E-05	0.3	0.06	0.6	LOC339975
rs116134219	4	188326029	0.96	1.03E-05	0.3	0.05	0.6	LOC339975
rs113165089	4	188327733	0.95	1.13E-05	0.3	0.06	0.6	LOC339975
rs112232951	4	188329318	0.96	1.09E-05	0.3	0.05	0.6	LOC339975
rs76911486	4	188333134	0.96	2.34E-05	0.4	0.07	0.6	LOC339975
rs75828282	4	188333577	0.96	2.40E-05	0.4	0.07	0.6	LOC339975
rs78465996	4	188338304	0.96	1.40E-05	0.3	0.07	0.6	LOC339975
rs113266307	4	188338566	0.96	1.39E-05	0.3	0.07	0.6	LOC339975
rs11933754	4	188339771	0.96	1.38E-05	0.3	0.08	0.6	LOC339975
rs11935240	4	188341193	0.96	1.42E-05	0.3	0.07	0.6	LOC339975
rs144115853	4	188341998	0.96	1.31E-05	0.3	0.08	0.6	LOC339975
rs77414185	4	188344450	0.97	1.29E-05	0.3	0.07	0.6	LOC339975
rs77097179	4	188344799	0.97	1.39E-05	0.3	0.06	0.6	LOC339975
rs75536828	4	188357518	1.00	8.23E-06	0.3	0.14	0.6	LOC339975
rs57438669	4	188360032	0.99	4.02E-06	0.2	0.08	0.6	LOC339975

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs76655731	4	188360163	1.00	8.30E-06	0.3	0.09	0.6	LOC339975
rs74464601	4	188360229	1.00	6.02E-06	0.2	0.09	0.6	LOC339975
rs11931480	4	188362893	1.00	6.15E-06	0.2	0.09	0.6	LOC339975
rs77547664	4	188371961	0.99	4.61E-06	0.2	0.07	0.6	LOC339975
rs76656386	4	188375913	0.99	6.19E-06	0.2	0.07	0.6	LOC339975
rs76446531	4	188379086	0.99	6.20E-06	0.2	0.07	0.6	LOC339975
rs75496321	4	188381891	0.99	4.82E-06	0.2	0.07	0.6	LOC339975
rs112168621	4	188383350	0.99	6.62E-06	0.2	0.06	0.6	LOC339975
rs77667574	4	188391705	1.00	5.40E-06	0.2	0.08	0.6	LOC339975
rs4862808	4	188398633	1.00	8.28E-06	0.3	0.08	0.6	LOC339975
rs4862809	4	188398663	1.00	1.14E-05	0.3	0.10	0.6	LOC339975
rs76307063	4	188398832	1.00	4.77E-06	0.2	0.08	0.6	LOC339975
rs79440070	4	188399004	1.00	4.75E-06	0.2	0.08	0.6	LOC339975
rs13133236	4	188399321	1.00	8.13E-06	0.3	0.08	0.6	LOC339975
rs13133469	4	188399430	1.00	8.11E-06	0.3	0.08	0.6	LOC339975
rs4862812	4	188400337	1.00	7.96E-06	0.3	0.08	0.6	LOC339975
rs4862813	4	188400742	1.00	8.00E-06	0.3	0.08	0.6	LOC339975
rs79071854	4	188403196	1.00	3.89E-06	0.2	0.09	0.6	LOC339975
rs112858899	4	188408330	1.00	3.24E-06	0.2	0.09	0.6	LOC339975
rs143715208	4	188409336	1.00	3.23E-06	0.2	0.09	0.6	LOC339975
rs147704297	4	188409464	1.00	3.23E-06	0.2	0.09	0.6	LOC339975
rs113574664	4	188411134	1.00	3.20E-06	0.2	0.09	0.6	LOC339975
rs113237987	4	188411190	1.00	3.20E-06	0.2	0.09	0.6	LOC339975
rs111557585	4	188411234	1.00	3.21E-06	0.2	0.07	0.6	LOC339975
rs116733741	4	188411375	1.00	3.20E-06	0.2	0.07	0.6	LOC339975
rs75061382	4	188412397	1.00	3.19E-06	0.2	0.07	0.6	LOC339975
rs76107485	4	188414534	1.00	3.17E-06	0.2	0.07	0.6	LOC339975
rs77477912	4	188416674	1.00	3.16E-06	0.2	0.09	0.6	LOC339975
rs77867835	4	188417458	1.00	3.16E-06	0.2	0.09	0.6	LOC339975
rs111755541	4	188419930	1.00	3.16E-06	0.2	0.09	0.6	LOC339975
rs1367555	4	188420909	1.00	5.54E-06	0.2	0.07	0.6	LOC339975
rs77692574	4	188421215	1.00	3.16E-06	0.2	0.09	0.6	LOC339975
rs10001779	4	188423355	1.00	5.54E-06	0.2	0.07	0.6	LOC339975
rs9996952	4	188424578	1.00	7.97E-06	0.3	0.06	0.6	LOC339975
rs112048494	4	188425380	1.00	3.50E-06	0.2	0.08	0.6	LOC339975
rs11726136	4	188426585	0.96	1.05E-06	0.2	0.09	0.6	LOC339975
rs6841040	4	188426755	0.94	3.17E-06	0.2	0.07	0.6	LOC339975
rs61300960	4	188428337	0.99	5.99E-06	0.2	0.06	0.6	intergenic
rs75525298	4	188428338	0.99	4.00E-06	0.2	0.09	0.6	intergenic
rs10026638	4	188429196	0.97	2.50E-06	0.2	0.13	0.6	intergenic
rs11736878	4	188431095	0.97	1.21E-06	0.2	0.14	0.6	intergenic
rs10032833	4	188431411	0.97	1.73E-06	0.2	0.11	0.6	intergenic
rs11929993	4	188432233	0.96	3.36E-05	0.5	0.22	0.7	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs11930080	4	188432452	0.99	8.22E-06	0.3	0.07	0.6	intergenic
rs9996885	4	188433623	0.99	5.88E-06	0.2	0.07	0.6	intergenic
rs28475332	4	188434677	0.96	1.91E-06	0.2	0.10	0.6	intergenic
rs28579034	4	188434935	0.96	2.80E-06	0.2	0.12	0.6	intergenic
rs77598188	4	188435868	0.98	4.34E-06	0.2	0.09	0.6	intergenic
rs145889915	4	188442359	0.97	4.80E-06	0.2	0.09	0.6	intergenic
rs28876409	4	188443338	0.96	8.20E-06	0.3	0.07	0.6	intergenic
rs112567828	4	188443982	0.96	6.18E-06	0.2	0.07	0.6	intergenic
rs145753563	4	188445993	0.94	8.52E-06	0.3	0.07	0.6	intergenic
rs140865211	5	1486075	0.72	2.41E-05	0.4	0.25	0.7	*LPCAT1*
rs186220053	5	65346295	0.68	1.21E-05	0.3	0.55	0.9	*ERBB2IP*
rs1561398	5	71486785	1.00	2.81E-05	0.4	0.51	0.9	*MAP1B*
rs10077261	5	77880332	0.91	2.85E-05	0.5	0.85	1.0	*LHFPL2*
rs10474560	5	77880436	0.91	2.84E-05	0.5	0.86	1.0	*LHFPL2*
rs9293756	5	77918136	1.00	2.31E-05	0.4	0.37	0.8	*LHFPL2*
rs6868848	5	77928880	0.90	1.88E-05	0.4	0.12	0.6	*LHFPL2*
rs6868849	5	77928881	0.90	1.74E-05	0.4	0.11	0.6	*LHFPL2*
rs61757417	5	142816370	0.90	1.92E-05	0.4	0.64	1.0	*NR3C1*
rs2546694	5	168960626	0.58	6.42E-06	0.2	0.15	0.6	intergenic
rs150080552	5	175673833	0.86	2.31E-05	0.4	0.63	1.0	*C5orf25*
rs140024910	5	175677335	0.87	1.80E-05	0.4	0.68	1.0	*C5orf25*
rs75753052	5	175680389	0.87	1.18E-05	0.3	0.69	1.0	*C5orf25*
rs6863942	5	175682713	0.88	1.52E-05	0.3	0.70	1.0	*C5orf25*
rs56145408	5	175684714	0.88	1.47E-05	0.3	0.76	1.0	*C5orf25*
rs148614682	5	175687581	0.89	2.20E-05	0.4	0.78	1.0	*C5orf25*
rs189483775	5	175688102	0.89	1.49E-05	0.3	0.87	1.0	*C5orf25*
rs7736390	5	175691922	0.89	1.49E-05	0.3	0.86	1.0	*C5orf25*
rs113708040	5	175699526	0.91	2.37E-05	0.4	0.96	1.0	*C5orf25*
rs150930492	5	175704361	0.80	3.11E-05	0.5	0.21	0.7	*C5orf25*
rs71599477	5	175704389	0.80	7.62E-07	0.2	0.63	1.0	*C5orf25*
rs140753856	5	175704589	0.90	1.60E-05	0.4	0.93	1.0	*C5orf25*
rs140351658	5	175706000	0.90	1.24E-05	0.3	0.89	1.0	*C5orf25*
rs149533417	5	175706083	0.90	1.60E-05	0.4	0.89	1.0	*C5orf25*
rs141966194	5	175707578	0.89	2.42E-05	0.4	0.84	1.0	*C5orf25*
rs142491407	5	175715202	0.91	2.08E-05	0.4	0.83	1.0	*C5orf25*
rs56164224	5	175719092	0.94	3.76E-05	0.5	0.96	1.0	*C5orf25*
rs6860403	5	175734589	0.99	4.44E-05	0.5	0.91	1.0	*C5orf25*
rs66505320	5	175737851	0.99	4.44E-05	0.5	0.91	1.0	*C5orf25*
rs66532964	5	175793171	0.98	4.48E-05	0.5	0.87	1.0	*ARL10*
rs67175907	5	175795653	0.96	4.30E-05	0.5	0.83	1.0	*ARL10*
rs78999839	6	1531294	0.88	3.94E-06	0.2	NA	NA	intergenic
rs114403119	6	1532809	0.89	5.95E-06	0.2	NA	NA	intergenic
rs141397728	6	1533251	0.88	3.90E-06	0.2	NA	NA	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs77034495	6	1548032	0.88	1.17E-05	0.3	NA	NA	intergenic
rs76107504	6	14252112	0.62	8.54E-06	0.3	0.36	0.8	intergenic
rs145978565	6	14324540	0.60	3.88E-05	0.5	NA	NA	intergenic
rs62398911	6	20506272	0.80	3.11E-05	0.5	0.42	0.9	intergenic
rs1737236	6	22350806	1.00	3.06E-05	0.5	0.16	0.6	intergenic
rs1737235	6	22351177	1.00	3.06E-05	0.5	0.16	0.6	intergenic
rs1772075	6	22351325	1.00	2.93E-05	0.5	0.16	0.6	intergenic
rs1737232	6	22351919	1.00	3.46E-05	0.5	0.16	0.6	intergenic
rs1737231	6	22352142	1.00	3.84E-05	0.5	0.16	0.6	intergenic
rs1737230	6	22352366	1.00	3.62E-05	0.5	0.16	0.6	intergenic
rs1737229	6	22352455	1.00	3.62E-05	0.5	0.16	0.6	intergenic
rs1737225	6	22353202	1.00	3.62E-05	0.5	0.16	0.6	intergenic
rs1629125	6	22353431	1.00	3.62E-05	0.5	0.16	0.6	intergenic
rs1772077	6	22353749	1.00	3.62E-05	0.5	0.16	0.6	intergenic
rs1772078	6	22353844	1.00	3.64E-05	0.5	0.16	0.6	intergenic
rs1772079	6	22354105	1.00	3.63E-05	0.5	0.16	0.6	intergenic
rs1772080	6	22354220	1.00	3.64E-05	0.5	0.16	0.6	intergenic
rs1772081	6	22354351	1.00	3.64E-05	0.5	0.16	0.6	intergenic
rs1772082	6	22354566	1.00	3.64E-05	0.5	0.16	0.6	intergenic
rs1737224	6	22355309	1.00	2.48E-06	0.2	0.19	0.6	intergenic
rs1772083	6	22355537	1.00	2.51E-06	0.2	0.20	0.6	intergenic
rs1772084	6	22355624	0.99	4.31E-06	0.2	0.17	0.6	intergenic
rs1772085	6	22355889	0.97	3.07E-05	0.5	0.28	0.7	intergenic
rs1737223	6	22355948	0.98	8.23E-06	0.3	0.23	0.7	intergenic
rs9366446	6	22356438	0.98	2.23E-06	0.2	0.15	0.6	intergenic
rs9379384	6	22356478	0.97	8.44E-07	0.2	0.19	0.6	intergenic
rs66815324	6	22356528	0.97	2.24E-06	0.2	0.37	0.8	intergenic
rs113931720	6	22356614	0.99	2.73E-06	0.2	0.26	0.7	intergenic
rs111776589	6	22356621	0.96	1.67E-05	0.4	0.35	0.8	intergenic
rs113397700	6	22356664	0.98	1.61E-06	0.2	0.22	0.7	intergenic
rs111410562	6	22356677	0.97	7.41E-06	0.3	0.12	0.6	intergenic
rs78526335	6	22356703	0.93	2.40E-05	0.4	0.69	1.0	intergenic
rs73393321	6	22356753	0.97	3.41E-06	0.2	0.46	0.9	intergenic
rs1205944	6	22356823	1.00	2.48E-06	0.2	0.19	0.6	intergenic
rs1205943	6	22356914	1.00	2.50E-06	0.2	0.21	0.7	intergenic
rs34183148	6	22356970	0.95	1.67E-05	0.4	0.36	0.8	intergenic
rs35363281	6	22356972	0.94	1.14E-05	0.3	0.29	0.7	intergenic
rs35097027	6	22356983	0.94	7.34E-06	0.3	0.40	0.8	intergenic
rs35519411	6	22356987	0.95	1.06E-05	0.3	0.36	0.8	intergenic
rs1205942	6	22357011	0.95	1.06E-05	0.3	0.36	0.8	intergenic
rs1205941	6	22357033	0.95	4.06E-05	0.5	0.33	0.8	intergenic
rs1205940	6	22357113	0.98	1.06E-05	0.3	0.26	0.7	intergenic
rs1205939	6	22357130	0.99	2.03E-06	0.2	0.18	0.6	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs1205938	6	22357296	1.00	2.47E-06	0.2	0.19	0.6	intergenic
rs1318602	6	22357342	1.00	2.42E-06	0.2	0.21	0.6	intergenic
rs1318601	6	22357345	1.00	2.47E-06	0.2	0.19	0.6	intergenic
rs1205937	6	22357993	1.00	2.88E-06	0.2	0.16	0.6	intergenic
rs1205936	6	22358116	1.00	2.48E-06	0.2	0.19	0.6	intergenic
rs1205935	6	22358129	1.00	2.48E-06	0.2	0.19	0.6	intergenic
rs1205934	6	22358196	1.00	2.48E-06	0.2	0.19	0.6	intergenic
rs1205933	6	22358361	1.00	2.48E-06	0.2	0.19	0.6	intergenic
rs1205932	6	22358391	1.00	2.48E-06	0.2	0.18	0.6	intergenic
rs1205931	6	22358551	1.00	2.43E-06	0.2	0.21	0.6	intergenic
rs1205930	6	22359007	1.00	2.88E-06	0.2	0.16	0.6	intergenic
rs1205929	6	22359126	1.00	2.64E-06	0.2	0.19	0.6	intergenic
rs1205928	6	22359465	1.00	3.00E-06	0.2	0.16	0.6	intergenic
rs1205927	6	22359577	1.00	2.43E-06	0.2	0.19	0.6	intergenic
rs1205926	6	22359759	1.00	2.43E-06	0.2	0.19	0.6	intergenic
rs714589	6	22359969	1.00	1.82E-06	0.2	0.20	0.6	intergenic
rs714590	6	22360029	1.00	2.43E-06	0.2	0.19	0.6	intergenic
rs714591	6	22360046	1.00	2.43E-06	0.2	0.19	0.6	intergenic
rs1008200	6	22360741	1.00	2.35E-06	0.2	0.18	0.6	intergenic
rs1205925	6	22361476	1.00	2.85E-06	0.2	0.21	0.7	intergenic
rs1205924	6	22361929	1.00	2.83E-06	0.2	0.16	0.6	intergenic
rs1205923	6	22362760	1.00	2.43E-06	0.2	0.15	0.6	intergenic
rs1205921	6	22364446	1.00	1.92E-06	0.2	0.20	0.6	intergenic
rs1205919	6	22365242	1.00	1.44E-06	0.2	0.15	0.6	intergenic
rs1205916	6	22367062	1.00	1.34E-06	0.2	0.16	0.6	intergenic
rs1205913	6	22371494	0.99	2.35E-06	0.2	0.13	0.6	intergenic
rs1100578	6	22371990	0.99	1.20E-06	0.2	0.18	0.6	intergenic
rs1100576	6	22373216	0.98	2.04E-06	0.2	0.11	0.6	intergenic
rs1100575	6	22373710	0.97	2.41E-06	0.2	0.11	0.6	intergenic
rs1100583	6	22375081	0.97	2.20E-06	0.2	0.11	0.6	intergenic
rs138132573	6	32978261	1.00	3.58E-06	0.2	NA	NA	*HLA-DOA*
rs146683088	6	32984983	1.00	3.57E-06	0.2	NA	NA	intergenic
rs150406963	6	32989951	0.99	3.57E-06	0.2	NA	NA	intergenic
rs114104288	6	32997592	0.99	4.11E-06	0.2	NA	NA	intergenic
rs181275409	6	32999504	0.99	3.74E-06	0.2	NA	NA	intergenic
rs181330926	6	33004097	1.00	3.58E-06	0.2	NA	NA	intergenic
rs115338659	6	33004745	0.98	8.99E-06	0.3	NA	NA	intergenic
rs150480104	6	33005434	1.00	3.73E-06	0.2	NA	NA	intergenic
rs188001307	6	33008064	1.00	3.58E-06	0.2	NA	NA	intergenic
rs116519378	6	33017238	0.99	1.83E-06	0.2	NA	NA	intergenic
rs116832467	6	33017950	1.00	2.20E-06	0.2	NA	NA	intergenic
rs143148648	6	49908506	0.85	2.25E-05	0.4	0.85	1.0	intergenic
rs239777	6	54830038	0.78	3.73E-05	0.5	0.05	0.6	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs12209224	6	77030438	0.81	4.47E-05	0.5	0.31	0.8	intergenic
rs2485801	6	93585436	1.00	3.89E-05	0.5	0.51	0.9	intergenic
rs2138340	6	93586078	1.00	3.76E-05	0.5	0.51	0.9	intergenic
rs2506909	6	93597512	1.00	3.55E-05	0.5	0.51	0.9	intergenic
rs1572632	6	93598512	1.00	3.54E-05	0.5	0.51	0.9	intergenic
rs16870432	6	93638655	0.90	2.46E-05	0.4	0.96	1.0	intergenic
rs188106941	6	104905706	0.94	1.44E-05	0.3	0.90	1.0	intergenic
rs111329411	6	114611601	0.70	3.68E-05	0.5	0.44	0.9	intergenic
rs2356078	6	118567605	1.00	3.12E-05	0.5	0.95	1.0	*SLC35F1*
rs281871	6	118571414	1.00	3.12E-05	0.5	0.95	1.0	*SLC35F1*
rs1387918	6	129663271	1.00	3.96E-05	0.5	0.90	1.0	*LAMA2*
rs7766742	6	129666380	1.00	1.67E-05	0.4	0.95	1.0	*LAMA2*
rs9398900	6	129669931	1.00	1.56E-05	0.4	0.99	1.0	*LAMA2*
rs13193413	6	129677895	0.95	4.21E-05	0.5	0.99	1.0	*LAMA2*
rs9372926	6	129684622	0.99	2.29E-05	0.4	0.88	1.0	*LAMA2*
rs63196263	6	129687026	0.99	2.47E-05	0.4	0.95	1.0	*LAMA2*
rs13194587	6	129694064	0.98	2.01E-05	0.4	0.84	1.0	*LAMA2*
rs4298368	6	129694961	0.97	2.25E-05	0.4	0.92	1.0	*LAMA2*
rs9388700	6	129697797	0.98	2.82E-05	0.4	0.83	1.0	*LAMA2*
rs79898774	6	137477313	0.57	3.72E-06	0.2	0.92	1.0	*IL22RA2*
rs6923972	6	138909661	0.99	3.95E-05	0.5	0.65	1.0	intergenic
rs75350744	6	143647820	0.94	3.21E-05	0.5	0.63	1.0	*AIG1*
rs77919378	6	143655982	0.94	2.22E-05	0.4	0.67	1.0	*AIG1*
rs78754824	6	143657178	0.94	2.23E-05	0.4	0.67	1.0	*AIG1*
rs76184032	6	143662887	0.89	4.14E-05	0.5	0.77	1.0	intergenic
rs6930238	6	145228890	0.99	4.21E-05	0.5	0.43	0.9	intergenic
rs75556273	6	147688190	1.00	4.21E-05	0.5	0.47	0.9	*STXBP5*
rs73787171	6	147689457	1.00	4.18E-05	0.5	0.42	0.9	*STXBP5*
rs117826317	6	147691033	1.00	4.21E-05	0.5	0.47	0.9	*STXBP5*
rs117646839	6	147694579	1.00	4.21E-05	0.5	0.47	0.9	*STXBP5*
rs79806067	6	147695718	0.94	1.59E-05	0.4	0.55	0.9	*STXBP5*
rs78647861	6	147695746	1.00	4.21E-05	0.5	0.47	0.9	*STXBP5*
rs9497762	6	147696197	1.00	4.18E-05	0.5	0.43	0.9	*STXBP5*
rs1034245	6	147706791	1.00	4.19E-05	0.5	0.43	0.9	*STXBP5*
rs12194637	6	151263284	0.94	2.62E-05	0.4	0.06	0.6	*MTHFD1L*
rs803448	6	151264132	0.98	4.48E-05	0.5	0.05	0.6	*MTHFD1L*
rs45619636	6	166823764	0.62	5.75E-07	0.1	0.66	1.0	*RPS6KA2*
rs145686314	7	1386960	0.60	3.39E-05	0.5	0.49	0.9	intergenic
rs113147352	7	1637024	0.57	3.19E-06	0.2	0.53	0.9	intergenic
rs143614258	7	1696933	0.60	2.23E-05	0.4	NA	NA	intergenic
rs118120700	7	5721117	0.68	7.42E-06	0.3	0.79	1.0	*RNF216*
rs116997033	7	5837618	0.62	3.60E-05	0.5	0.91	1.0	intergenic
rs145944844	7	9386123	0.78	1.94E-05	0.4	0.91	1.0	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs9769459	7	47786445	1.00	1.09E-05	0.3	0.93	1.0	intergenic
rs62457904	7	55580097	0.93	1.56E-05	0.4	NA	NA	*VOPP1*
rs149460737	7	113494910	0.94	2.26E-05	0.4	0.58	0.9	intergenic
rs138373005	7	113607020	0.99	3.72E-05	0.5	0.93	1.0	intergenic
rs144790732	7	118685098	0.39	1.22E-05	0.3	0.58	1.0	intergenic
rs2170478	8	3464175	0.99	1.07E-05	0.3	0.06	0.6	*CSMD1*
rs2469338	8	3466917	1.00	1.84E-05	0.4	0.03	0.6	*CSMD1*
rs2469334	8	3467396	0.82	1.04E-05	0.3	0.14	0.6	*CSMD1*
rs2469331	8	3467762	0.99	1.67E-05	0.4	0.03	0.6	*CSMD1*
rs2469329	8	3468180	1.00	2.68E-05	0.4	0.04	0.6	*CSMD1*
rs1121619	8	3469700	0.99	2.67E-05	0.4	0.04	0.6	*CSMD1*
rs72639918	8	36200105	0.88	3.64E-05	0.5	0.08	0.6	intergenic
rs143420877	8	51696312	0.85	3.02E-05	0.5	0.76	1.0	*SNTG1*
rs72645876	8	51720873	0.88	4.04E-05	0.5	0.84	1.0	intergenic
rs149203983	8	51727143	0.88	4.15E-05	0.5	0.85	1.0	intergenic
rs10087464	8	51763094	0.99	4.18E-05	0.5	0.69	1.0	intergenic
rs7846269	8	51831461	0.96	3.39E-05	0.5	0.80	1.0	intergenic
rs117535834	8	76073376	0.60	4.27E-05	0.5	0.57	0.9	intergenic
rs557807	8	90615624	0.99	2.76E-05	0.4	0.63	1.0	intergenic
rs79048468	8	137895174	0.89	1.74E-07	0.1	0.53	0.9	intergenic
rs117687198	8	137907672	0.90	6.78E-08	0.1	0.40	0.8	intergenic
rs148161159	8	137910630	0.89	1.27E-05	0.3	0.09	0.6	intergenic
rs75683128	8	142506570	0.37	1.70E-05	0.4	0.53	0.9	*FLJ43860*
rs149489509	8	144360947	0.53	2.17E-05	0.4	0.06	0.6	intergenic
rs3750416	9	2054380	0.96	5.72E-06	0.2	0.78	1.0	*SMARCA2*
rs79825622	9	2058984	0.97	4.16E-06	0.2	0.76	1.0	*SMARCA2*
rs113524072	9	2069929	0.99	1.05E-05	0.3	NA	NA	*SMARCA2*
rs78301395	9	2071686	0.99	1.12E-05	0.3	NA	NA	*SMARCA2*
rs12342240	9	2074021	0.98	7.42E-06	0.3	0.53	0.9	*SMARCA2*
rs17471283	9	2271886	0.81	3.72E-05	0.5	0.76	1.0	intergenic
rs2589305	9	23558910	0.80	1.43E-06	0.2	0.30	0.7	intergenic
rs12380156	9	28007018	1.00	1.18E-05	0.3	0.58	0.9	*LINGO2*
rs10968272	9	28007239	1.00	1.18E-05	0.3	0.57	0.9	*LINGO2*
rs10968273	9	28007813	1.00	1.18E-05	0.3	0.58	0.9	*LINGO2*
rs79144456	9	28008378	1.00	1.18E-05	0.3	0.58	1.0	*LINGO2*
rs117562410	9	28008569	1.00	1.18E-05	0.3	0.58	1.0	*LINGO2*
rs182385046	9	70993692	0.43	1.09E-05	0.3	0.27	0.7	*PGM5*
rs138328727	9	75953590	0.76	1.84E-05	0.4	0.43	0.9	intergenic
rs142071081	9	76024441	0.75	3.58E-05	0.5	0.75	1.0	intergenic
rs182666170	9	98325920	0.65	1.64E-05	0.4	NA	NA	intergenic
rs138160086	9	118602789	0.96	4.44E-05	0.5	0.76	1.0	intergenic
rs139178463	9	121440368	0.93	1.47E-05	0.3	0.61	1.0	intergenic
rs140651197	9	121440507	0.92	1.56E-05	0.4	0.88	1.0	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs1335261	9	121453610	1.00	3.51E-05	0.5	0.83	1.0	intergenic
rs7045014	9	136658734	0.77	3.24E-05	0.5	0.50	0.9	*VAV2*
rs73553919	9	136659327	0.86	4.12E-05	0.5	0.50	0.9	*VAV2*
rs59232037	9	136664901	0.90	4.12E-05	0.5	0.62	1.0	*VAV2*
rs59158070	9	136665094	0.90	4.11E-05	0.5	0.62	1.0	*VAV2*
rs73663847	9	136666574	0.90	3.94E-05	0.5	0.62	1.0	*VAV2*
rs1548380	9	136668393	0.91	3.80E-05	0.5	0.62	1.0	*VAV2*
rs56035119	9	136671469	0.94	2.77E-05	0.4	0.49	0.9	*VAV2*
rs55940550	9	136672358	0.92	3.46E-05	0.5	0.70	1.0	*VAV2*
rs28578536	9	136673248	0.94	2.61E-05	0.4	0.50	0.9	*VAV2*
rs7042276	9	136675739	0.95	2.58E-05	0.4	0.51	0.9	*VAV2*
rs2077392	9	136679457	0.96	2.84E-05	0.5	0.55	0.9	*VAV2*
rs7047672	9	136679921	1.00	2.54E-05	0.4	0.56	0.9	*VAV2*
rs78132252	10	70984526	0.72	1.48E-05	0.3	0.03	0.6	*HKDC1*
rs189208264	10	76878339	0.55	1.96E-05	0.4	0.71	1.0	*SAMD8*
rs2670223	10	80127055	0.96	4.12E-05	0.5	0.26	0.7	intergenic
rs150976965	10	81525997	0.56	2.12E-05	0.4	0.85	1.0	intergenic
rs191720898	10	100860117	0.80	1.40E-05	0.3	NA	NA	*HPSE2*
rs12784542	10	102769710	0.95	3.13E-05	0.5	0.06	0.6	*PDZD7*
rs10509835	10	109205283	0.98	4.51E-05	0.5	0.28	0.7	intergenic
rs147541483	10	109628602	0.44	2.03E-05	0.4	0.24	0.7	intergenic
rs2420781	10	122370247	0.92	4.08E-05	0.5	0.63	1.0	intergenic
rs150067672	10	123915941	0.71	1.51E-05	0.3	0.83	1.0	*TACC2*
rs76537500	10	130053296	0.78	2.52E-05	0.4	0.81	1.0	intergenic
rs11597322	10	131492126	0.99	2.59E-05	0.4	0.74	1.0	*MGMT*
rs4751111	10	131502995	1.00	9.56E-06	0.3	0.55	0.9	*MGMT*
rs11815846	10	131504196	1.00	8.34E-06	0.3	0.90	1.0	*MGMT*
rs11016884	10	131504495	1.00	4.92E-06	0.2	0.78	1.0	*MGMT*
rs78037285	10	134367856	0.86	3.72E-05	0.5	0.78	1.0	*INPP5A*
rs141710236	10	134400013	0.85	3.81E-05	0.5	0.90	1.0	*INPP5A*
rs149345061	10	134447849	0.82	1.42E-05	0.3	0.89	1.0	*INPP5A*
rs150499775	10	134495006	0.70	4.25E-05	0.5	0.79	1.0	*INPP5A*
rs117644802	10	134498721	0.69	3.56E-05	0.5	0.79	1.0	*INPP5A*
rs114202853	10	134517256	0.75	1.28E-05	0.3	0.46	0.9	*INPP5A*
rs61873115	11	2318964	0.75	4.28E-05	0.5	0.36	0.8	*C11orf21*
rs4929975	11	2319303	0.75	4.29E-05	0.5	0.37	0.8	*C11orf21*
rs71488794	11	21841125	0.59	3.35E-05	0.5	NA	NA	intergenic
rs188553612	11	22050642	0.56	3.99E-05	0.5	0.63	1.0	intergenic
rs117023318	11	22240250	0.60	3.93E-05	0.5	0.90	1.0	*ANO5*
rs72882668	11	31062562	0.64	2.25E-06	0.2	NA	NA	LOC1005087
rs72888131	11	31286032	0.89	5.52E-06	0.2	NA	NA	*DCDC1*
rs149522059	11	36980477	0.61	1.37E-05	0.3	0.78	1.0	intergenic
rs117496632	11	36983355	0.62	1.79E-05	0.4	0.47	0.9	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs145100044	11	67569004	0.59	3.34E-05	0.5	0.59	1.0	*FAM86C2P*
rs76380214	11	68149280	0.70	3.97E-06	0.2	0.95	1.0	*LRP5*
rs141570922	11	68244370	0.89	3.86E-06	0.2	0.56	0.9	*PPP6R3*
rs181741343	11	68245257	0.91	6.11E-06	0.2	0.59	1.0	*PPP6R3*
rs185193806	11	68404877	0.66	7.61E-06	0.3	0.93	1.0	intergenic
rs191238062	11	109396279	0.54	3.55E-05	0.5	0.70	1.0	intergenic
rs11218131	11	120928722	1.00	4.42E-05	0.5	0.73	1.0	*TBCEL*
rs78610152	11	129354006	0.89	1.24E-05	0.3	0.84	1.0	intergenic
rs3019806	11	129374559	0.97	2.88E-05	0.5	0.71	1.0	intergenic
rs7942352	11	129380953	0.94	1.35E-05	0.3	0.77	1.0	intergenic
rs7941324	11	129384263	0.93	2.63E-05	0.4	0.62	1.0	intergenic
rs113240933	11	132910459	0.63	3.93E-05	0.5	0.96	1.0	*OPCML*
rs11047037	12	23825407	0.74	4.08E-05	0.5	0.98	1.0	*SOX5*
rs17176610	12	65114763	1.00	6.13E-06	0.2	0.48	0.9	*GNS*
rs79073680	12	65156596	0.96	1.03E-05	0.3	0.71	1.0	intergenic
rs17764405	12	65202269	0.89	2.81E-05	0.4	0.50	0.9	intergenic
rs17176841	12	65204874	0.95	3.48E-05	0.5	0.89	1.0	intergenic
rs79280005	12	84181090	0.62	3.10E-05	0.5	0.89	1.0	intergenic
rs73207068	12	88446859	0.73	9.23E-06	0.3	NA	NA	*CEP290*
rs182110748	13	34484961	0.83	2.18E-05	0.4	NA	NA	*RFC3*
rs75689743	13	34487594	0.83	2.20E-05	0.4	NA	NA	*RFC3*
rs146733701	13	39814716	0.82	2.46E-05	0.4	NA	NA	intergenic
rs111596958	13	69770858	0.99	3.92E-05	0.5	0.96	1.0	intergenic
rs77304322	13	69772371	0.99	3.84E-05	0.5	0.96	1.0	intergenic
rs117695261	13	74325506	0.57	1.16E-07	0.1	0.82	1.0	*KLF12*
rs17377290	13	85167241	0.76	4.31E-05	0.5	0.06	0.6	intergenic
rs146736842	13	85187435	0.81	3.84E-05	0.5	0.05	0.6	intergenic
rs189306812	13	87150672	0.67	4.45E-05	0.5	0.25	0.7	intergenic
rs76003828	13	88044777	0.85	8.30E-06	0.3	0.61	1.0	intergenic
rs9513263	13	88051360	0.84	7.25E-06	0.3	0.84	1.0	intergenic
rs149154304	13	92479378	0.79	6.59E-06	0.2	0.94	1.0	*GPC5*
rs7982492	13	93543743	0.96	3.19E-05	0.5	0.15	0.6	intergenic
rs113211811	13	106903793	0.88	3.03E-05	0.5	0.31	0.8	intergenic
rs142025631	14	28311095	0.79	3.41E-05	0.5	NA	NA	intergenic
rs72712459	14	65734600	0.97	2.08E-05	0.4	0.19	0.6	intergenic
rs117727648	14	77100157	0.54	3.36E-08	0.1	0.89	1.0	intergenic
rs192688395	14	77211718	0.55	2.70E-07	0.1	1.00	1.0	intergenic
rs113653607	14	77641605	0.80	3.61E-07	0.1	0.39	0.8	intergenic
rs142687658	14	77750946	0.94	1.47E-07	0.1	NA	NA	*POMT2*
rs150915059	14	77755799	0.93	6.61E-07	0.2	NA	NA	*POMT2*
rs150017190	14	77806492	0.94	3.47E-07	0.1	NA	NA	intergenic
rs56198483	14	77828430	0.85	1.53E-07	0.1	0.08	0.6	*TMED8*
rs115653266	14	77843222	0.78	3.91E-06	0.2	NA	NA	*TMED8*

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs145730217	14	77940220	0.92	1.41E-06	0.2	NA	NA	intergenic
rs142877874	14	77975928	0.88	2.62E-06	0.2	NA	NA	*SPTLC2*
rs141530129	14	77994346	0.81	1.61E-05	0.4	NA	NA	*SPTLC2*
rs139545749	14	78078188	0.91	2.99E-06	0.2	NA	NA	*SPTLC2*
rs6574877	14	86695129	1.00	2.61E-05	0.4	0.01	0.6	intergenic
rs1505173	14	86695795	1.00	2.67E-05	0.4	0.01	0.6	intergenic
rs150334896	14	86696062	0.99	2.80E-05	0.4	0.01	0.55	intergenic
rs187136545	14	86775771	0.64	7.63E-07	0.2	NA	NA	intergenic
rs74919671	14	87140828	0.81	4.89E-06	0.2	0.15	0.6	intergenic
rs17128291	14	92882826	1.00	1.17E-05	0.3	0.15	0.6	*SLC24A4*
rs7158887	14	94241933	0.79	2.10E-05	0.4	1.00	1.0	*PRIMA1*
rs35980137	14	94245799	0.86	9.77E-07	0.2	0.73	1.0	*PRIMA1*
rs75051938	14	103459051	0.90	4.39E-05	0.5	0.56	0.9	*CDC42BPB*
rs2269326	14	103476438	0.83	9.01E-06	0.3	0.50	0.9	*CDC42BPB*
rs2285017	14	103484440	0.82	7.53E-06	0.3	0.52	0.9	*CDC42BPB*
rs35485315	14	103503352	0.81	2.06E-05	0.4	0.42	0.9	*CDC42BPB*
rs62010548	15	33946424	0.87	6.50E-06	0.2	0.18	0.6	*RYR3*
rs7163100	15	33951138	0.98	2.33E-05	0.4	0.44	0.9	*RYR3*
rs7178736	15	33951141	0.98	2.16E-05	0.4	0.38	0.8	*RYR3*
rs28735770	15	33959197	0.96	1.83E-05	0.4	0.67	1.0	*RYR3*
rs78828848	15	33968766	0.85	2.15E-06	0.2	0.03	0.6	*RYR3*
rs4780153	15	33972420	1.00	1.69E-07	0.1	0.95	1.0	*RYR3*
rs12907278	15	33973177	1.00	2.31E-06	0.2	0.65	1.0	*RYR3*
rs2172855	15	33977340	0.98	1.58E-06	0.2	0.62	1.0	*RYR3*
rs141657190	15	33977672	0.99	2.09E-06	0.2	0.65	1.0	*RYR3*
rs11631255	15	33978895	0.99	5.02E-06	0.2	0.60	1.0	*RYR3*
rs62012577	15	33979644	0.86	9.65E-06	0.3	0.21	0.7	*RYR3*
rs28602988	15	33982726	0.99	3.53E-05	0.5	0.51	0.9	*RYR3*
rs7169195	15	33983296	0.99	3.58E-05	0.5	0.51	0.9	*RYR3*
rs2088143	15	33983545	0.99	4.14E-05	0.5	0.44	0.9	*RYR3*
rs7174044	15	33983566	0.99	4.23E-05	0.5	0.46	0.9	*RYR3*
rs1353348	15	33984848	1.00	3.19E-07	0.1	0.76	1.0	*RYR3*
rs11632363	15	33984987	1.00	3.95E-06	0.2	0.63	1.0	*RYR3*
rs12592542	15	33985818	1.00	3.89E-06	0.2	0.63	1.0	*RYR3*
rs939432	15	33986294	0.99	2.42E-07	0.1	0.81	1.0	*RYR3*
rs2076954	15	33987341	1.00	2.84E-06	0.2	0.85	1.0	*RYR3*
rs2339298	15	33987505	0.99	3.50E-05	0.5	0.30	0.7	*RYR3*
rs6495216	15	33994836	1.00	1.60E-05	0.4	0.67	1.0	*RYR3*
rs11072621	15	33995495	1.00	2.37E-05	0.4	0.77	1.0	*RYR3*
rs8040310	15	34005613	1.00	4.50E-05	0.5	0.93	1.0	*RYR3*
rs118013228	15	50274508	0.89	3.43E-05	0.5	0.35	0.8	*ATP8B4*
rs111511606	15	50346721	0.96	4.89E-06	0.2	0.37	0.8	*ATP8B4*
rs113002981	15	50349114	0.99	8.01E-06	0.3	0.81	1.0	*ATP8B4*

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs72742523	15	79713106	0.60	3.77E-07	0.1	0.94	1.0	intergenic
rs182003445	15	82246207	0.80	2.98E-05	0.5	0.25	0.7	intergenic
rs113550094	15	84248208	0.83	4.44E-06	0.2	0.67	1.0	*SH3GL3*
rs7183304	15	92420700	0.83	4.52E-07	0.1	0.79	1.0	*SLCO3A1*
rs62029493	15	94313625	0.65	2.98E-05	0.5	0.48	0.9	intergenic
rs547673	16	1113621	0.51	1.52E-05	0.3	0.27	0.7	intergenic
rs562647	16	1117662	0.54	1.63E-05	0.4	0.13	0.6	LOC146336
rs669264	16	1120645	0.54	4.38E-05	0.5	0.14	0.6	LOC146336
rs77730974	16	11800685	0.82	3.50E-05	0.5	NA	NA	*TXNDC11*
rs76512865	16	11924051	0.97	4.10E-05	0.5	0.76	1.0	intergenic
rs7184515	16	12001447	1.00	4.53E-05	0.5	0.73	1.0	*GSPT1*
rs13335494	16	12001772	0.98	2.84E-05	0.5	0.68	1.0	*GSPT1*
rs4781162	16	12003035	1.00	4.52E-05	0.5	0.73	1.0	*GSPT1*
rs28880613	16	12018374	0.96	4.20E-05	0.5	0.65	1.0	intergenic
rs117060081	16	50245502	0.79	1.71E-05	0.4	0.91	1.0	*PAPD5*
rs116988464	16	50355115	0.79	6.55E-06	0.2	0.58	1.0	*BRD7*
rs8058581	16	66361269	0.86	2.41E-05	0.4	0.45	0.9	intergenic
rs12599734	16	73451181	0.95	1.13E-05	0.3	0.21	0.7	LOC1005061
rs35614540	16	73458459	0.97	3.24E-06	0.2	0.25	0.7	intergenic
rs35236088	16	73458610	0.97	3.15E-06	0.2	0.22	0.7	intergenic
rs6564189	16	73460072	0.96	3.77E-05	0.5	0.22	0.7	intergenic
rs141652736	16	82538255	0.82	5.67E-06	0.2	0.96	1.0	intergenic
rs16954993	17	5616349	0.98	3.95E-05	0.5	0.02	0.6	intergenic
rs11867305	17	5622186	0.98	3.74E-05	0.5	0.02	0.6	intergenic
rs77142476	17	7032762	0.53	4.53E-05	0.5	0.93	1.0	intergenic
rs141874876	17	36949890	0.70	4.30E-06	0.2	NA	NA	*PIP4K2B*
rs111885856	17	50788068	0.68	1.62E-05	0.4	0.79	1.0	intergenic
rs147319530	18	30888675	0.55	1.24E-05	0.3	NA	NA	*C18orf34*
rs73470849	18	40174449	0.99	4.47E-05	0.5	0.67	1.0	LOC284260
rs73470851	18	40174503	0.99	4.41E-05	0.5	0.67	1.0	LOC284260
rs55868360	18	40174704	0.99	4.46E-05	0.5	0.67	1.0	LOC284260
rs56300458	18	40174748	0.99	4.17E-05	0.5	0.67	1.0	LOC284260
rs1594434	18	40174924	0.99	3.99E-05	0.5	0.67	1.0	LOC284260
rs8088979	18	40176118	1.00	6.57E-06	0.2	0.82	1.0	LOC284260
rs10048404	18	54578482	1.00	1.89E-05	0.4	0.68	1.0	*WDR7*
rs117779952	18	56698860	0.91	3.51E-05	0.5	0.79	1.0	intergenic
rs77201620	18	63158485	0.71	4.07E-05	0.5	0.60	1.0	intergenic
rs60388875	19	3339134	0.60	5.49E-06	0.2	0.60	1.0	intergenic
rs773916	19	16856864	1.00	1.59E-05	0.4	0.78	1.0	*NWD1*
rs186607440	19	29649259	0.56	2.99E-06	0.2	0.95	1.0	intergenic
rs71356041	19	37703129	0.93	4.29E-05	0.5	0.25	0.7	intergenic
rs2161520	19	37945039	0.98	8.15E-06	0.3	0.08	0.6	*ZNF569*
rs77379396	19	38169881	1.00	4.23E-05	0.5	0.85	1.0	*ZNF781*

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs112555037	19	38171107	1.00	4.23E-05	0.5	0.85	1.0	*ZNF781*
rs74774310	19	38173433	1.00	4.23E-05	0.5	0.85	1.0	*ZNF781*
rs3095726	19	38229824	0.92	4.42E-05	0.5	0.09	0.6	*ZNF573*
rs144953949	19	40726691	0.77	4.45E-05	0.5	0.88	1.0	intergenic
rs3730053	19	40743739	0.92	4.63E-06	0.2	0.91	1.0	*AKT2*
rs76137255	19	40783832	0.92	4.27E-06	0.2	0.94	1.0	*AKT2*
rs190665219	19	40889198	0.74	3.54E-05	0.5	0.26	0.7	*HIPK4*
rs1861792	19	46558503	0.97	8.22E-06	0.3	0.97	1.0	intergenic
rs2160735	19	46558752	0.97	5.07E-06	0.2	0.48	0.9	intergenic
rs4802285	19	46559146	0.98	2.80E-05	0.4	0.83	1.0	intergenic
rs8111075	19	46560510	0.99	7.99E-06	0.3	0.93	1.0	intergenic
rs2009468	19	46560881	0.99	3.20E-05	0.5	0.75	1.0	intergenic
rs2216348	19	46561785	1.00	2.89E-05	0.5	0.75	1.0	intergenic
rs4802287	19	46561988	1.00	2.82E-05	0.4	0.76	1.0	intergenic
rs11669138	19	46562524	0.99	6.71E-06	0.2	0.89	1.0	intergenic
rs8101164	19	46563471	1.00	2.49E-05	0.4	0.76	1.0	LOC400706
rs8101445	19	46563653	1.00	2.36E-05	0.4	0.76	1.0	LOC400706
rs7250943	19	46564525	1.00	2.10E-05	0.4	0.80	1.0	LOC400706
rs2880179	19	46564972	1.00	2.05E-05	0.4	0.80	1.0	LOC400706
rs4803895	19	46565481	0.99	3.03E-06	0.2	0.49	0.9	LOC400706
rs1035254	19	46566661	1.00	2.04E-05	0.4	0.80	1.0	LOC400706
rs16980258	19	46566808	1.00	1.98E-05	0.4	0.80	1.0	LOC400706
rs35237775	19	46566881	1.00	2.05E-05	0.4	0.81	1.0	LOC400706
rs11083798	19	46568638	1.00	1.74E-05	0.4	0.73	1.0	LOC400706
rs713409	19	46569619	1.00	2.17E-05	0.4	0.75	1.0	LOC400706
rs713411	19	46569630	1.00	2.17E-05	0.4	0.74	1.0	LOC400706
rs713412	19	46569831	0.99	2.21E-05	0.4	0.73	1.0	LOC400706
rs7251500	19	46571411	0.97	7.44E-06	0.3	0.76	1.0	LOC400706
rs7250786	19	46571668	0.97	1.20E-05	0.3	0.79	1.0	LOC400706
rs4802288	19	46590131	0.90	1.24E-05	0.3	0.54	0.9	intergenic
rs2341902	19	46601339	0.90	1.22E-05	0.3	0.56	0.9	intergenic
rs8102694	19	46602690	0.90	1.21E-05	0.3	0.56	0.9	intergenic
rs2080618	19	46608823	0.87	1.93E-05	0.4	0.48	0.9	intergenic
rs10418330	19	46611282	0.82	2.47E-05	0.4	0.25	0.7	intergenic
rs145409557	20	2849618	0.86	3.82E-05	0.5	NA	NA	*PTPRA*
rs59330923	20	8198472	1.00	3.45E-05	0.5	0.53	0.9	*PLCB1*
rs2876128	20	8205124	1.00	2.06E-05	0.4	0.50	0.9	*PLCB1*
rs6055677	20	8210961	1.00	4.49E-05	0.5	0.51	0.9	*PLCB1*
rs6055678	20	8211200	1.00	4.12E-05	0.5	0.52	0.9	*PLCB1*
rs6108126	20	8211226	1.00	4.36E-05	0.5	0.50	0.9	*PLCB1*
rs6055679	20	8211285	1.00	4.36E-05	0.5	0.47	0.9	*PLCB1*
rs6055680	20	8211559	1.00	4.36E-05	0.5	0.50	0.9	*PLCB1*
rs181084932	20	53517832	0.42	3.80E-05	0.5	0.07	0.6	intergenic

A.3. Association results for AD in the IASPSAD with *q*-values ≤ 0.5 and COGA replication

SNP	CHR	BP	INFO	IASPSAD *p*	IASPSAD *q*	COGA *p*	COGA *q*	GENE
rs2268260	21	35941233	0.84	1.04E-05	0.3	0.82	1.0	*RCAN1*
rs140130045	21	38017102	0.91	1.67E-05	0.4	0.07	0.6	intergenic
rs112814695	21	38022514	0.92	4.40E-05	0.5	0.06	0.6	intergenic
rs113277403	21	38029653	0.95	3.81E-05	0.5	0.18	0.6	intergenic
rs113916689	21	38037118	0.80	2.92E-05	0.5	0.02	0.6	intergenic
rs4819991	22	17750681	0.68	9.90E-06	0.3	0.92	1.0	intergenic
rs62236482	22	17751803	0.64	1.41E-05	0.3	0.89	1.0	intergenic
rs80270647	22	17752392	0.63	1.36E-05	0.3	0.89	1.0	intergenic
rs75748085	22	17759713	0.57	2.94E-05	0.5	0.50	0.9	intergenic
rs5749005	22	17759885	0.57	2.99E-05	0.5	0.50	0.9	intergenic
rs9613388	22	21305473	0.68	2.18E-05	0.4	0.26	0.7	*CRKL*
rs118082458	22	35222253	0.93	4.38E-05	0.5	0.42	0.9	intergenic
rs62243094	22	35380612	0.97	4.25E-05	0.5	0.29	0.7	intergenic
rs9610187	22	35381350	0.97	4.22E-05	0.5	0.29	0.7	intergenic
rs62243095	22	35381546	0.97	4.22E-05	0.5	0.29	0.7	intergenic
rs9610188	22	35382016	0.98	4.23E-05	0.5	0.28	0.7	intergenic
rs62243096	22	35387936	0.99	2.98E-05	0.5	0.36	0.8	intergenic
rs713985	22	35388018	0.99	2.98E-05	0.5	0.36	0.8	intergenic
rs9607217	22	35388905	1.00	3.83E-05	0.5	0.36	0.8	intergenic
rs9610190	22	35388938	1.00	3.63E-05	0.5	0.36	0.8	intergenic
rs9607218	22	35389220	1.00	3.63E-05	0.5	0.36	0.8	intergenic
rs7284941	22	35389727	1.00	3.63E-05	0.5	0.36	0.8	intergenic
rs7286439	22	35390051	1.00	3.63E-05	0.5	0.36	0.8	intergenic
rs7285304	22	35390101	1.00	3.61E-05	0.5	0.36	0.8	intergenic
rs62243111	22	35390127	1.00	3.61E-05	0.5	0.37	0.8	intergenic
rs9610191	22	35390306	1.00	3.61E-05	0.5	0.36	0.8	intergenic
rs9607219	22	35390327	1.00	3.64E-05	0.5	0.36	0.8	intergenic
rs9610192	22	35390669	1.00	3.80E-05	0.5	0.36	0.8	intergenic
rs6518917	22	35390846	1.00	3.89E-05	0.5	0.36	0.8	intergenic
rs6518918	22	35391073	1.00	4.05E-05	0.5	0.36	0.8	intergenic
rs6518920	22	35391265	1.00	4.19E-05	0.5	0.36	0.8	intergenic
rs142314105	22	41667156	0.88	1.53E-05	0.4	0.68	1.0	*RANGAP1*
rs116434743	22	41783206	0.91	3.33E-05	0.5	0.91	1.0	*TEF*
rs117530017	22	41785067	0.91	3.30E-05	0.5	0.99	1.0	*TEF*
rs138772877	22	41822725	0.93	2.63E-05	0.4	0.80	1.0	intergenic
rs185631468	22	42710609	0.85	6.56E-08	0.1	NA	NA	intergenic
rs59649586	22	48652075	0.40	7.67E-06	0.3	0.66	1.0	intergenic

A.4. Association results for the joint meta-analysis of AD in the IASPSAD and COGA

SNP	CHR	BP	Z	META *p*	DIR	META *q*	COGA *p*	IASPSAD *p*	GENE
rs114720027	1	1226690	4.68	2.82E-06	++	0.3	3.27E-03	2.64E-04	*SCNN1D*
rs11590002	1	175032518	-4.41	1.04E-05	--	0.5	2.98E-01	2.17E-06	intergenic
rs11582546	1	175049009	5.06	4.23E-07	++	0.3	3.04E-02	2.58E-06	*TNN*
rs61827435	1	175066967	5.13	2.87E-07	++	0.3	2.31E-02	2.44E-06	*TNN*
rs114901102	1	210901331	-4.45	8.75E-06	--	0.4	1.34E-01	9.06E-06	*KCNH1*
rs1002159	2	23538632	4.48	7.43E-06	++	0.4	3.28E-02	6.27E-05	intergenic
rs10203610	2	27572265	4.59	4.35E-06	++	0.3	5.40E-02	1.71E-05	*GTF3C2*
rs13011459	2	166012871	-4.53	5.91E-06	--	0.3	8.02E-03	2.39E-04	*SCN3A*
rs73011180	3	3256506	4.63	3.75E-06	++	0.3	2.47E-03	4.47E-04	intergenic
rs143582116	3	3300915	4.54	5.56E-06	++	0.3	1.91E-03	8.13E-04	intergenic
rs73149932	3	147209249	-4.45	8.71E-06	--	0.4	3.94E-01	8.39E-07	intergenic
rs150268941	3	172575312	5.30	1.16E-07	++	0.3	2.02E-01	1.39E-08	intergenic
rs9870004	3	192625680	4.37	1.25E-05	++	0.5	2.25E-06	5.65E-02	*MB21D2*
rs11719444	3	192626348	4.40	1.09E-05	++	0.5	2.07E-06	5.33E-02	*MB21D2*
rs9853424	3	192629549	-4.53	5.90E-06	--	0.3	1.59E-06	3.98E-02	*MB21D2*
rs114613162	4	58610828	-4.64	3.51E-06	--	0.3	2.93E-03	3.62E-04	intergenic
rs113600173	4	58649957	4.72	2.42E-06	++	0.3	5.42E-02	8.47E-06	intergenic
rs1911588	4	148021144	-4.42	9.79E-06	--	0.4	9.72E-04	2.28E-03	intergenic
rs6823293	4	188314906	-4.74	2.19E-06	--	0.3	2.63E-02	2.10E-05	LOC339975
rs78616487	4	188322087	-4.60	4.29E-06	--	0.3	5.07E-02	1.84E-05	LOC339975
rs11943588	4	188322659	4.62	3.84E-06	++	0.3	6.20E-02	1.20E-05	LOC339975
rs116134219	4	188326029	-4.69	2.78E-06	--	0.3	5.32E-02	1.03E-05	LOC339975
rs113165089	4	188327733	-4.64	3.43E-06	--	0.3	5.91E-02	1.13E-05	LOC339975
rs112232951	4	188329318	4.68	2.94E-06	++	0.3	5.34E-02	1.09E-05	LOC339975
rs76911486	4	188333134	4.46	8.04E-06	++	0.4	7.11E-02	2.34E-05	LOC339975
rs75828282	4	188333577	-4.46	8.23E-06	--	0.4	7.13E-02	2.40E-05	LOC339975
rs78465996	4	188338304	-4.55	5.31E-06	--	0.3	7.20E-02	1.40E-05	LOC339975
rs113266307	4	188338566	-4.55	5.30E-06	--	0.3	7.21E-02	1.39E-05	LOC339975
rs11933754	4	188339771	4.53	5.84E-06	++	0.3	7.82E-02	1.38E-05	LOC339975
rs11935240	4	188341193	4.54	5.53E-06	++	0.3	7.34E-02	1.42E-05	LOC339975
rs144115853	4	188341998	-4.54	5.71E-06	--	0.3	7.93E-02	1.31E-05	LOC339975
rs77414185	4	188344450	-4.56	5.18E-06	--	0.3	7.44E-02	1.29E-05	LOC339975
rs77097179	4	188344799	4.59	4.42E-06	++	0.3	6.27E-02	1.39E-05	LOC339975
rs75536828	4	188357518	4.46	8.17E-06	++	0.4	1.35E-01	8.23E-06	LOC339975
rs57438669	4	188360032	4.72	2.35E-06	++	0.3	8.44E-02	4.02E-06	LOC339975
rs76655731	4	188360163	-4.59	4.50E-06	--	0.3	8.79E-02	8.30E-06	LOC339975
rs74464601	4	188360229	4.64	3.46E-06	++	0.3	8.79E-02	6.02E-06	LOC339975
rs11931480	4	188362893	4.63	3.64E-06	++	0.3	9.00E-02	6.15E-06	LOC339975
rs77547664	4	188371961	4.74	2.12E-06	++	0.3	7.21E-02	4.61E-06	LOC339975
rs76656386	4	188375913	-4.70	2.65E-06	--	0.3	7.11E-02	6.19E-06	LOC339975
rs76446531	4	188379086	4.69	2.68E-06	++	0.3	7.15E-02	6.20E-06	LOC339975
rs75496321	4	188381891	4.75	2.03E-06	++	0.3	6.78E-02	4.82E-06	LOC339975

A.4. Association results for the joint meta-analysis of AD in the IASPSAD and COGA

SNP	CHR	BP	Z	META *p*	DIR	META *q*	COGA *p*	IASPSAD *p*	GENE
rs112168621	4	188383350	-4.72	2.34E-06	--	0.3	6.19E-02	6.62E-06	LOC339975
rs77667574	4	188391705	4.68	2.86E-06	++	0.3	8.17E-02	5.40E-06	LOC339975
rs4862808	4	188398633	-4.61	3.98E-06	--	0.3	8.04E-02	8.28E-06	LOC339975
rs4862809	4	188398663	-4.51	6.61E-06	--	0.4	9.66E-02	1.14E-05	LOC339975
rs76307063	4	188398832	-4.71	2.44E-06	--	0.3	7.83E-02	4.77E-06	LOC339975
rs79440070	4	188399004	4.71	2.43E-06	++	0.3	7.83E-02	4.75E-06	LOC339975
rs13133236	4	188399321	-4.62	3.91E-06	--	0.3	8.03E-02	8.13E-06	LOC339975
rs13133469	4	188399430	-4.62	3.91E-06	--	0.3	8.03E-02	8.11E-06	LOC339975
rs4862812	4	188400337	4.63	3.64E-06	++	0.3	7.71E-02	7.96E-06	LOC339975
rs4862813	4	188400742	-4.63	3.58E-06	--	0.3	7.59E-02	8.00E-06	LOC339975
rs79071854	4	188403196	4.71	2.50E-06	++	0.3	9.00E-02	3.89E-06	LOC339975
rs112858899	4	188408330	4.75	2.09E-06	++	0.3	8.80E-02	3.24E-06	LOC339975
rs143715208	4	188409336	-4.75	2.08E-06	--	0.3	8.80E-02	3.23E-06	LOC339975
rs147704297	4	188409464	-4.75	2.08E-06	--	0.3	8.80E-02	3.23E-06	LOC339975
rs113574664	4	188411134	4.75	2.07E-06	++	0.3	8.80E-02	3.20E-06	LOC339975
rs113237987	4	188411190	4.75	2.07E-06	++	0.3	8.80E-02	3.20E-06	LOC339975
rs111557585	4	188411234	4.79	1.64E-06	++	0.3	7.43E-02	3.21E-06	LOC339975
rs116733741	4	188411375	-4.79	1.64E-06	--	0.3	7.43E-02	3.20E-06	LOC339975
rs75061382	4	188412397	4.79	1.63E-06	++	0.3	7.43E-02	3.19E-06	LOC339975
rs76107485	4	188414534	4.80	1.62E-06	++	0.3	7.43E-02	3.17E-06	LOC339975
rs77477912	4	188416674	4.75	2.05E-06	++	0.3	8.80E-02	3.16E-06	LOC339975
rs77867835	4	188417458	-4.75	2.05E-06	--	0.3	8.80E-02	3.16E-06	LOC339975
rs111755541	4	188419930	4.75	2.05E-06	++	0.3	8.80E-02	3.16E-06	LOC339975
rs1367555	4	188420909	4.71	2.51E-06	++	0.3	7.30E-02	5.54E-06	LOC339975
rs77692574	4	188421215	-4.75	2.05E-06	--	0.3	8.79E-02	3.16E-06	LOC339975
rs10001779	4	188423355	4.71	2.54E-06	++	0.3	7.38E-02	5.54E-06	LOC339975
rs9996952	4	188424578	-4.72	2.40E-06	--	0.3	5.60E-02	7.97E-06	LOC339975
rs112048494	4	188425380	4.75	2.01E-06	++	0.3	8.17E-02	3.50E-06	LOC339975
rs11726136	4	188426585	-4.92	8.66E-07	--	0.3	9.00E-02	1.05E-06	LOC339975
rs6841040	4	188426755	4.80	1.60E-06	++	0.3	7.36E-02	3.17E-06	LOC339975
rs61300960	4	188428337	-4.74	2.09E-06	--	0.3	6.07E-02	5.99E-06	intergenic
rs75525298	4	188428338	-4.71	2.45E-06	--	0.3	8.72E-02	4.00E-06	intergenic
rs10026638	4	188429196	4.68	2.94E-06	++	0.3	1.29E-01	2.50E-06	intergenic
rs11736878	4	188431095	4.77	1.81E-06	++	0.3	1.36E-01	1.21E-06	intergenic
rs10032833	4	188431411	4.78	1.74E-06	++	0.3	1.10E-01	1.73E-06	intergenic
rs11930080	4	188432452	4.64	3.42E-06	++	0.3	7.21E-02	8.22E-06	intergenic
rs9996885	4	188433623	4.71	2.53E-06	++	0.3	7.09E-02	5.88E-06	intergenic
rs28475332	4	188434677	-4.78	1.73E-06	--	0.3	1.04E-01	1.91E-06	intergenic
rs28579034	4	188434935	-4.67	3.07E-06	--	0.3	1.25E-01	2.80E-06	intergenic
rs77598188	4	188435868	4.68	2.91E-06	++	0.3	9.41E-02	4.34E-06	intergenic
rs145889915	4	188442359	4.67	2.99E-06	++	0.3	9.04E-02	4.80E-06	intergenic
rs28876409	4	188443338	4.67	2.98E-06	++	0.3	6.51E-02	8.20E-06	intergenic

A.4. Association results for the joint meta-analysis of AD in the IASPSAD and COGA

SNP	CHR	BP	Z	META *p*	DIR	META *q*	COGA *p*	IASPSAD *p*	GENE
rs112567828	4	188443982	4.69	2.72E-06	++	0.3	7.25E-02	6.18E-06	intergenic
rs145753563	4	188445993	4.66	3.20E-06	++	0.3	6.70E-02	8.52E-06	intergenic
rs74973212	5	22730877	4.61	4.01E-06	++	0.3	7.91E-03	1.62E-04	*CDH12*
rs77454938	5	22892433	-4.57	4.90E-06	--	0.3	3.19E-04	2.62E-03	intergenic
rs189681803	5	22900277	-4.59	4.42E-06	--	0.3	3.57E-04	2.23E-03	intergenic
rs9716627	5	22933007	4.37	1.23E-05	++	0.5	9.12E-04	2.92E-03	intergenic
rs79708954	5	22964609	4.89	9.93E-07	++	0.3	1.76E-04	9.68E-04	intergenic
rs7728036	5	22970127	4.93	8.33E-07	++	0.3	1.59E-04	8.89E-04	intergenic
rs189537064	5	23187921	4.45	8.78E-06	++	0.4	3.17E-02	7.91E-05	intergenic
rs143975413	5	23187926	-4.45	8.78E-06	--	0.4	3.17E-02	7.91E-05	intergenic
rs111997333	5	148298840	-4.41	1.03E-05	--	0.5	2.80E-02	1.10E-04	intergenic
rs13232384	7	22186176	4.42	1.01E-05	++	0.5	9.97E-03	3.35E-04	*RAPGEF5*
rs12592	7	22202117	4.81	1.48E-06	++	0.3	9.50E-03	4.63E-05	*RAPGEF5*
rs1981601	7	22202485	-4.58	4.72E-06	--	0.3	1.38E-02	1.07E-04	*RAPGEF5*
rs2170478	8	3464175	-4.67	3.04E-06	--	0.3	5.54E-02	1.07E-05	*CSMD1*
rs2469338	8	3466917	-4.76	1.96E-06	--	0.3	2.64E-02	1.84E-05	*CSMD1*
rs2469334	8	3467396	4.41	1.04E-05	++	0.5	1.41E-01	1.04E-05	*CSMD1*
rs2469331	8	3467762	4.78	1.72E-06	++	0.3	2.53E-02	1.67E-05	*CSMD1*
rs2469329	8	3468180	4.62	3.78E-06	++	0.3	3.50E-02	2.68E-05	*CSMD1*
rs1121619	8	3469700	4.60	4.26E-06	++	0.3	3.87E-02	2.67E-05	*CSMD1*
rs144690006	8	107766355	-4.68	2.89E-06	--	0.3	1.38E-04	2.83E-03	intergenic
rs79048468	8	137895174	-4.56	5.22E-06	--	0.3	5.25E-01	1.74E-07	intergenic
rs117687198	8	137907672	4.81	1.49E-06	++	0.3	4.05E-01	6.78E-08	intergenic
rs148161159	8	137910630	4.51	6.51E-06	++	0.4	8.93E-02	1.27E-05	intergenic
rs149489509	8	144360947	4.54	5.65E-06	++	0.3	5.65E-02	2.17E-05	intergenic
rs78132252	10	70984526	-4.80	1.58E-06	--	0.3	2.59E-02	1.48E-05	*HKDC1*
rs12784542	10	102769710	-4.44	9.02E-06	--	0.4	6.41E-02	3.13E-05	*PDZD7*
rs78145908	11	131034265	4.40	1.07E-05	++	0.5	8.11E-04	2.79E-03	intergenic
rs142465076	12	70447861	-4.53	5.96E-06	--	0.3	1.08E-03	1.34E-03	intergenic
rs73188186	12	107846289	4.38	1.19E-05	++	0.5	2.01E-02	1.90E-04	*BTBD11*
rs138166981	12	107846404	4.38	1.17E-05	++	0.5	1.98E-02	1.90E-04	*BTBD11*
rs149633740	12	107846410	-4.38	1.17E-05	--	0.5	1.98E-02	1.89E-04	*BTBD11*
rs11113334	12	107847283	4.41	1.03E-05	++	0.5	1.74E-02	1.92E-04	*BTBD11*
rs189083626	12	121693090	-4.43	9.49E-06	--	0.4	3.13E-03	9.05E-04	*CAMKK2*
rs12306064	12	127603048	4.48	7.65E-06	++	0.4	1.75E-05	1.79E-02	intergenic
rs1039792	12	127606132	4.47	7.98E-06	++	0.4	1.93E-05	1.77E-02	intergenic
rs11058993	12	127606502	-4.47	7.96E-06	--	0.4	1.92E-05	1.77E-02	intergenic
rs61543828	12	127607255	-4.47	7.93E-06	--	0.4	1.92E-05	1.77E-02	intergenic
rs11613288	12	127611661	-4.56	5.18E-06	--	0.3	1.99E-05	1.27E-02	intergenic
rs34224343	12	127612167	4.64	3.52E-06	++	0.3	1.75E-05	1.01E-02	intergenic
rs34404953	12	127612361	4.64	3.50E-06	++	0.3	1.74E-05	1.01E-02	intergenic
rs78518547	12	127613578	-4.65	3.40E-06	--	0.3	1.73E-05	9.93E-03	intergenic

A.4. Association results for the joint meta-analysis of AD in the IASPSAD and COGA

SNP	CHR	BP	Z	META *p*	DIR	META *q*	COGA *p*	IASPSAD *p*	GENE
rs191968104	12	127614398	4.48	7.59E-06	++	0.4	1.32E-05	2.02E-02	intergenic
rs181655177	12	127614399	4.48	7.58E-06	++	0.4	1.32E-05	2.02E-02	intergenic
rs74981202	12	127615159	-4.65	3.28E-06	--	0.3	1.73E-05	9.66E-03	intergenic
rs11058997	12	127615824	-4.66	3.23E-06	--	0.3	1.73E-05	9.55E-03	intergenic
rs79805024	12	127617105	-4.66	3.14E-06	--	0.3	1.72E-05	9.37E-03	intergenic
rs56780684	12	127618799	4.67	3.04E-06	++	0.3	1.72E-05	9.14E-03	intergenic
rs58333498	12	127618854	4.67	3.03E-06	++	0.3	1.72E-05	9.13E-03	intergenic
rs11058998	12	127620311	-4.67	2.96E-06	--	0.3	1.72E-05	8.97E-03	intergenic
rs11058999	12	127620476	-4.67	2.96E-06	--	0.3	1.72E-05	8.97E-03	intergenic
rs11503010	12	127622486	4.67	3.02E-06	++	0.3	1.73E-05	9.08E-03	intergenic
rs117348945	12	127623739	4.67	3.09E-06	++	0.3	1.75E-05	9.20E-03	intergenic
rs12302509	12	127625419	4.60	4.18E-06	++	0.3	1.40E-05	1.28E-02	intergenic
rs12312674	12	127627973	-4.68	2.94E-06	--	0.3	1.43E-05	9.76E-03	intergenic
rs10847280	12	127628993	-4.55	5.47E-06	--	0.3	4.74E-05	8.61E-03	intergenic
rs10847281	12	127629300	-4.68	2.82E-06	--	0.3	1.21E-05	1.03E-02	intergenic
rs11612470	12	127629492	4.69	2.80E-06	++	0.3	1.19E-05	1.03E-02	intergenic
rs10847282	12	127629984	-4.60	4.16E-06	--	0.3	1.20E-05	1.37E-02	intergenic
rs117695004	12	127630545	-4.53	6.01E-06	--	0.3	1.69E-05	1.53E-02	intergenic
rs117558206	12	127631620	-4.70	2.64E-06	--	0.3	1.13E-05	1.01E-02	intergenic
rs78954849	12	127632323	-4.70	2.57E-06	--	0.3	1.05E-05	1.02E-02	intergenic
rs76027692	12	127632352	-4.70	2.56E-06	--	0.3	1.05E-05	1.02E-02	intergenic
rs74596781	12	127634115	4.70	2.55E-06	++	0.3	1.02E-05	1.03E-02	intergenic
rs11059001	12	127634847	4.70	2.55E-06	++	0.3	1.01E-05	1.04E-02	intergenic
rs74398605	12	127635455	4.69	2.73E-06	++	0.3	1.00E-05	1.09E-02	intergenic
rs80334165	12	127635520	-4.69	2.76E-06	--	0.3	1.00E-05	1.10E-02	intergenic
rs10847286	12	127636509	-4.74	2.12E-06	--	0.3	4.27E-06	1.34E-02	intergenic
rs12313900	12	127637474	4.74	2.09E-06	++	0.3	4.21E-06	1.34E-02	intergenic
rs11059002	12	127637964	-4.70	2.61E-06	--	0.3	6.01E-06	1.33E-02	intergenic
rs9538694	13	61001708	4.60	4.14E-06	++	0.3	4.22E-05	7.38E-03	*TDRD3*
rs9528178	13	61220743	-4.86	1.18E-06	--	0.3	3.62E-04	6.79E-04	intergenic
rs17377290	13	85167241	-4.38	1.17E-05	--	0.5	6.35E-02	4.31E-05	intergenic
rs146736842	13	85187435	4.44	8.84E-06	++	0.4	5.49E-02	3.84E-05	intergenic
rs79862295	13	85204822	-4.39	1.16E-05	--	0.5	5.31E-02	5.52E-05	intergenic
rs117455152	13	85205057	-4.38	1.16E-05	--	0.5	5.31E-02	5.53E-05	intergenic
rs9520259	13	107633055	-4.57	4.89E-06	--	0.3	8.40E-03	1.87E-04	intergenic
rs147277902	14	61720703	4.75	2.08E-06	++	0.3	1.12E-02	5.55E-05	intergenic
rs117727648	14	77100157	4.49	7.05E-06	++	0.4	8.91E-01	3.36E-08	intergenic
rs56198483	14	77828430	5.24	1.63E-07	++	0.3	8.25E-02	1.53E-07	*TMED8*
rs74919671	14	87140828	-4.52	6.27E-06	--	0.4	1.49E-01	4.89E-06	intergenic
rs62010548	15	33946424	4.40	1.07E-05	++	0.5	1.83E-01	6.50E-06	*RYR3*
rs78828848	15	33968766	-5.07	3.98E-07	--	0.3	3.27E-02	2.15E-06	*RYR3*
rs2899683	15	62938169	-4.42	9.75E-06	--	0.4	6.00E-03	5.26E-04	intergenic

A.4. Association results for the joint meta-analysis of AD in the IASPSAD and COGA

SNP	CHR	BP	Z	META *p*	DIR	META *q*	COGA *p*	IASPSAD *p*	GENE
rs938979	15	62939440	4.63	3.59E-06	++	0.3	2.12E-03	4.89E-04	intergenic
rs8041641	15	62941808	4.38	1.17E-05	++	0.5	1.70E-02	2.25E-04	*TLN2*
rs6494336	15	62941810	-4.55	5.33E-06	--	0.3	7.73E-03	2.22E-04	*TLN2*
rs12903725	15	62942792	4.44	9.07E-06	++	0.4	9.46E-03	3.16E-04	*TLN2*
rs12438306	15	62942835	4.43	9.56E-06	++	0.4	1.12E-02	2.81E-04	*TLN2*
rs8039436	15	62944576	4.52	6.17E-06	++	0.4	4.58E-03	4.26E-04	*TLN2*
rs12916083	15	62945250	4.49	7.00E-06	++	0.4	7.35E-03	3.10E-04	*TLN2*
rs56376861	15	62947019	4.50	6.72E-06	++	0.4	7.19E-03	3.04E-04	*TLN2*
rs12591037	15	62948909	4.79	1.64E-06	++	0.3	2.77E-03	1.79E-04	*TLN2*
rs12593694	15	62951029	4.44	9.21E-06	++	0.4	2.69E-03	9.96E-04	*TLN2*
rs7167388	15	62959101	4.82	1.47E-06	++	0.3	1.24E-03	3.23E-04	*TLN2*
rs1574119	15	62962294	4.46	8.30E-06	++	0.4	3.93E-03	6.54E-04	*TLN2*
rs72742462	15	77371206	4.37	1.23E-05	++	0.5	1.36E-03	2.20E-03	intergenic
rs45455291	16	16232976	4.40	1.11E-05	++	0.5	4.47E-03	7.76E-04	*ABCC1*
rs113328089	16	16235366	4.39	1.13E-05	++	0.5	4.50E-03	7.90E-04	*ABCC1*
rs4781906	16	18013160	-4.42	9.76E-06	--	0.4	3.17E-03	9.21E-04	intergenic
rs6498759	16	18020295	4.46	8.32E-06	++	0.4	1.79E-03	1.25E-03	intergenic
rs2923133	16	57732437	-4.91	9.01E-07	--	0.3	5.79E-06	6.25E-03	*CCDC135*
rs28868312	16	73314930	4.37	1.22E-05	++	0.5	7.16E-04	3.42E-03	intergenic
rs28825739	16	73314935	4.77	1.88E-06	++	0.3	8.75E-04	5.41E-04	intergenic
rs149157093	18	7864399	-4.38	1.22E-05	--	0.5	1.23E-02	3.27E-04	*PTPRM*
rs6042781	20	14510635	-4.50	6.72E-06	--	0.4	1.22E-05	1.92E-02	*MACROD2*
rs58677849	20	48429508	4.50	6.71E-06	++	0.4	3.08E-05	1.25E-02	*SLC9A8*
rs140130045	21	38017102	4.52	6.15E-06	++	0.4	7.19E-02	1.67E-05	intergenic
rs112814695	21	38022514	4.40	1.10E-05	++	0.5	5.97E-02	4.40E-05	intergenic
rs78624947	21	38023820	4.37	1.24E-05	++	0.5	5.64E-02	5.46E-05	intergenic
rs75877632	21	38033949	4.37	1.22E-05	++	0.5	5.03E-02	6.28E-05	intergenic
rs113916689	21	38037118	4.69	2.75E-06	++	0.3	2.48E-02	2.92E-05	intergenic

Notes: DIR refers to the direction of effect. "++" and "--" indicate that the same allele increases and decreases risk, respectively.

A.5. Association results for the externalizing factor score in the IASPSAD

SNP	CHR	BP	INFO	*p*	*q*	GENE
rs34100349	1	166465517	0.83	2.50E-06	0.4	intergenic
rs11579746	1	166486813	0.85	4.27E-06	0.4	intergenic
rs12070599	1	166499803	0.71	3.75E-07	0.2	intergenic
rs10918508	1	166503560	0.72	1.18E-07	0.2	intergenic
rs12028217	1	238584540	0.75	4.14E-06	0.4	intergenic
rs75651477	2	5704362	0.53	2.12E-06	0.4	intergenic
rs149985535	2	29735999	0.73	4.59E-06	0.4	*ALK*
rs62170104	2	114532234	0.99	9.08E-07	0.3	intergenic
rs72946678	2	114532605	0.98	4.80E-07	0.2	intergenic
rs62170106	2	114536469	1.00	5.26E-07	0.2	intergenic
rs59123894	2	114536834	1.00	5.26E-07	0.2	intergenic
rs56884727	2	114536952	1.00	2.19E-07	0.2	intergenic
rs60221109	2	114537300	1.00	2.25E-07	0.2	intergenic
rs58511785	2	114538464	1.00	5.26E-07	0.2	intergenic
rs11685086	2	114540117	1.00	5.45E-07	0.2	intergenic
rs11692142	2	114540218	1.00	5.12E-07	0.2	intergenic
rs4849276	2	114542568	1.00	5.58E-07	0.2	intergenic
rs3870303	2	114548110	1.00	6.07E-07	0.2	intergenic
rs3982388	2	114548860	1.00	1.14E-06	0.3	intergenic
rs57489086	2	114550308	0.99	9.53E-07	0.3	intergenic
rs7567042	2	114555675	0.99	3.37E-07	0.2	intergenic
rs62170136	2	114556912	0.99	2.49E-07	0.2	intergenic
rs7587985	2	114559118	0.98	9.29E-08	0.2	intergenic
rs2900745	2	114563960	0.97	1.34E-07	0.2	intergenic
rs7588108	2	114566404	0.97	2.29E-07	0.2	intergenic
rs4292100	2	114571647	0.97	3.08E-07	0.2	intergenic
rs72835407	2	114572589	0.98	1.02E-06	0.3	intergenic
rs62170141	2	114572676	0.98	1.02E-06	0.3	intergenic
rs62170142	2	114572796	0.98	8.35E-07	0.3	intergenic
rs62170143	2	114572913	0.98	9.96E-07	0.3	intergenic
rs62170145	2	114576025	0.98	1.79E-06	0.4	intergenic
rs4450608	2	114576118	0.98	1.80E-06	0.4	intergenic
rs7589706	2	114576819	0.99	1.84E-06	0.4	intergenic
rs4389338	2	114577383	0.98	1.74E-06	0.4	intergenic
rs11903680	2	114577551	0.97	2.55E-06	0.4	intergenic
rs62170148	2	114577728	0.99	1.75E-06	0.4	intergenic
rs6715643	2	114578252	0.97	2.04E-06	0.4	intergenic
rs146366850	2	114579180	0.88	7.71E-07	0.3	intergenic
rs4848353	2	114580423	1.00	3.22E-06	0.4	intergenic

A.5. Association results for the externalizing factor score in the IASPSAD

SNP	CHR	BP	INFO	*p*	*q*	GENE
rs10172576	2	114581008	1.00	3.92E-06	0.4	intergenic
rs62170155	2	114582948	1.00	3.31E-06	0.4	intergenic
rs57106417	2	114583321	1.00	3.33E-06	0.4	intergenic
rs61050591	2	114583915	1.00	3.68E-06	0.4	intergenic
rs72948684	2	114584490	1.00	3.37E-06	0.4	intergenic
rs4411702	2	114584909	1.00	3.37E-06	0.4	intergenic
rs11676646	2	114585305	1.00	3.38E-06	0.4	intergenic
rs10185135	2	114586483	1.00	3.38E-06	0.4	intergenic
rs10209659	2	114586805	1.00	3.38E-06	0.4	intergenic
rs4849285	2	114587432	1.00	3.14E-06	0.4	intergenic
rs4849286	2	114587447	1.00	3.14E-06	0.4	intergenic
rs4849287	2	114587481	1.00	3.14E-06	0.4	intergenic
rs11685296	2	114588257	0.96	4.51E-07	0.2	intergenic
rs11675908	2	114588784	1.00	3.37E-06	0.4	LOC100506762
rs4849288	2	114589155	1.00	3.37E-06	0.4	LOC100506762
rs62170159	2	114589534	0.93	3.50E-06	0.4	LOC100506762
rs9646927	2	114590106	0.98	2.98E-06	0.4	LOC100506762
rs79238088	2	114590284	0.99	2.63E-06	0.4	LOC100506762
rs4849289	2	114591178	1.00	3.40E-06	0.4	LOC100506762
rs4849290	2	114591194	1.00	3.40E-06	0.4	LOC100506762
rs11691664	2	114592257	1.00	3.38E-06	0.4	LOC100506762
rs4849291	2	114594099	1.00	3.38E-06	0.4	LOC100506762
rs62171760	2	114594883	1.00	3.16E-06	0.4	LOC100506762
rs62171761	2	114595287	1.00	3.06E-06	0.4	LOC100506762
rs4571059	2	114599209	0.99	2.37E-06	0.4	LOC100506762
rs1028245	2	114601335	0.99	2.13E-06	0.4	LOC100506762
rs4241126	2	114603003	0.99	3.38E-06	0.4	LOC100506762
rs149679967	3	193096170	0.88	4.38E-06	0.4	*ATP13A5*
rs17356266	4	94527621	0.74	4.30E-06	0.4	*GRID2*
rs2973135	5	152555110	0.94	1.24E-06	0.3	intergenic
rs3849719	5	173248856	1.00	4.30E-06	0.4	intergenic
rs75850170	8	117417368	0.68	4.88E-06	0.4	intergenic
rs1410200	9	120981143	0.85	4.70E-06	0.4	intergenic
rs4836713	9	120998947	0.92	4.69E-06	0.4	intergenic
rs11185686	9	137134266	0.65	1.74E-06	0.4	intergenic
rs12235472	9	137134549	0.65	1.75E-06	0.4	intergenic
rs12237177	9	137134557	0.66	1.86E-06	0.4	intergenic
rs56168824	9	137135968	0.78	2.83E-06	0.4	intergenic
rs2399972	10	13557945	1.00	3.99E-06	0.4	intergenic

A.5. Association results for the externalizing factor score in the IASPSAD

SNP	CHR	BP	INFO	*p*	*q*	GENE
rs10752308	10	13558641	1.00	4.00E-06	0.4	intergenic
rs11258443	10	13583831	0.96	3.34E-06	0.4	intergenic
rs140886841	12	11652416	0.55	3.56E-06	0.4	intergenic
rs12184900	13	101522004	0.62	1.32E-07	0.2	intergenic
rs77419548	13	101525115	0.69	2.59E-06	0.4	intergenic
rs148336229	15	54594443	0.97	3.31E-06	0.4	*UNC13C*
rs75469002	15	54597980	0.99	2.64E-06	0.4	*UNC13C*
rs71474867	15	54598259	0.99	2.63E-06	0.4	*UNC13C*
rs71474868	15	54598286	0.99	2.63E-06	0.4	*UNC13C*
rs12915790	15	54599640	0.99	2.58E-06	0.4	*UNC13C*
rs12898437	15	54600325	0.99	2.78E-06	0.4	*UNC13C*
rs34432386	15	54600639	0.99	2.77E-06	0.4	*UNC13C*
rs8025641	15	54601303	0.99	2.77E-06	0.4	*UNC13C*
rs35926155	15	54604477	0.99	2.60E-06	0.4	*UNC13C*
rs12916554	15	54607255	1.00	2.50E-06	0.4	*UNC13C*
rs4404011	15	54997983	0.93	5.37E-06	0.5	intergenic
rs4889357	16	81688626	0.88	4.95E-07	0.2	*CMIP*

Vita

Laura Michele Hack was born on January 4th, 1984 in Columbia, MD. She graduated from Glenelg Country School in Columbia, MD in 2002 and from The College of William & Mary in Williamsbug, VA in 2006 *magna cum laude* with a Bachelor of Science in Neuroscience. As an undergraduate, she completed a Senior Honors thesis entitled "Thermoregulatory Neurons in the Anterior Hypothalamus". Following graduation, she entered the MD-PhD program at Virginia Commonwealth University in Richmond, VA. She has been in Richmond for the past six years, the first two of which were didactic medical training while the subsequent four have been spent conducting research at the Virginia Institute for Psychiatric and Behavioral Genetics. She was inducted into the Honor Society of Phi Kappa Phi in 2010 and has presented her research at the World Congress of Psychiatric Genetics in 2009 and 2011 as well as the American Physician Scientist's Association Annual Meeting in 2010. She will return to medical school for the remaining two clinical years in the Fall of 2012.