2019

Genome-Wide Systems Genetics of Alcohol Consumption and Dependence

Kristin Mignogna

GENOME-WIDE SYSTEMS GENETICS OF ALCOHOL CONSUMPTION AND DEPENDENCE

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctorate of Philosophy from Virginia Commonwealth University

By

KRISTIN M. MIGNOGNA

Directed by: MICHAEL F. MILES, MD, PhD

Professor, Departments of Pharmacology & Toxicology, Neurology, and Human & Molecular Genetics

Virginia Commonwealth University

Richmond, Virginia

May, 2019

Contributions

The work discussed in this manuscript involved the guidance and assistance of several other students, faculty, and staff members at Virginia Commonwealth University and collaborating institutes. Lorna Macleod, M.S., laboratory technician and manager, worked with Ms. Mignogna to collect behavioral data from the Diversity Outbred mice and their progenitor strains, described in Chapters 3-5 of this dissertation; and extracted the RNA from the Prefrontal Cortex of 220 of the Diversity Outbred mice, in preparation for the RNAseq analysis discussed in Chapter 5. Jessica Jurmain, B.A. also contributed substantially to the collection of the behavioral data analyzed in Chapters 3-5. Ms. Macleod, Ms. Jurmain, Andrew van der vaart, Ph.D., and Guy Harris, M.S. all contributed to the dissections and tissue collection from the Diversity Outbred mice. Christopher Pais, B.A. and Alexander Pais, B.A. collected the blood samples from the Diversity Outbred progenitor mice described in Chapter 3. Finally, Ria Shah (undergraduate Bioinformatics student at VCU) assisted in narrowing down lists of candidate genes by integrating data from the behavioral quantitative trait locus analysis with data from previous, relevant studies, as described in Chapter 4. Ms. Shah also performed functional overrepresentation analysis for candidate networks discussed in Chapter 5. All of these individuals were members of the Miles Laboratory in Virginia Commonwealth University's (VCU) Department of Pharmacology & Toxicology.

Jennifer Wolstenholme, Ph.D. (professor of Pharmacology and Toxicology at VCU) provided scientific advice regarding the interpretation and analysis of mouse behaviors studied in Chapters 3-5. Bradley Webb, Ph.D. and Brien Riley, Ph.D. (professors of Psychiatry in VCU's

Virginia Institute for Psychiatric, Behavioral, and Statistical Genetics (VIPBG)) advised on the genotypic quality control and analysis performed in Chapter 4 of this manuscript. Daniel Gatti, Ph.D. (Research Scientist at Jackson Laboratories), Benjamin Pejser, B.S. (Genomic Market Development Manager at Neogen GeneSeek Operations), and Andrew Morgan, Ph.D. (M.D. Ph.D. student at University of North Carolina, Chapel Hill) also provided advice and scripts for portions of the quality control performed in Chapter 4. Sliviu Bacanu, Ph.D. (professor of Psychiatry in VCU's VIPBG) provided guidance on the statistical analyses performed in all projects described in this dissertation. Amy Olex, M.S. (bioinformatician for VCU's Center for Clinical and Translational Research) and Mikhail Dozmorov, Ph.D. (VCU professor of Biostatistics) provided guidance and scripts for RNAseq data alignment, feature counting, gene expression quality control, and statistical analysis of gene expression data (all discussed in Chapter 5). Finally, Michael Miles, M.D., Ph.D. (VCU professor of Pharmacology & Toxicology, Neurology, and Human & Molecular Genetics) served as the primary advisor for the work performed for this dissertation.

All work described here was funded by the following grants and scholarships: National Institute for Alcoholism and Alcohol Abuse (NIAAA) P50 AA022537 (awarded to the VCU Alcohol Research Center), NIAAA P20 AA017828, NIAAA U01 AAO16667, and VCU Dissertation Assistantship Award.

Some text from this dissertation was taken directly from publications (both published in peer-reviewed journals, and in preparation) originally written by Kristin M. Mignogna, the author of this dissertation. These manuscripts are as follows:

Mignogna, K. M., Bacanu, S. A., Riley, B. P., Wolen, A. R., Miles, M. F. (2019). Cross-species alcohol dependence-associated gene networks: Co-analysis of mouse brain gene expression and human genome-wide association data. *Plos One*: *14*(4). doi: 10.1371/journal.pone.0202063

Mignogna, K. M., Macleod, L. M., Jurmain, J. L., Bacanu, S. A., Miles, M. F. (in preparation). Characterization and heritability of voluntary ethanol consumption and preference in the Diversity Outbred progenitor strains, under chronic and intermittent access.

Mignogna, K. M., Macleod, L. M., Riley, B. P., Webb, B. T., Shah, R. K., Miles, M. F. (in preparation) Behavioral quantitative trait loci for initial and long-term ethanol drinking behaviors in Diversity Outbred mice, under an intermittent access model.

Mignogna, K. M., Rodriguez, N. E., Macleod, L. M., Dozmorov, M. G., Olex, A. L., Miles, M. F. (in preparation). Ethanol drinking-associated gene expression patterns in the prefrontal cortex of voluntarily drinking Diversity Outbred mice.

Table of Contents

List of Abbreviations

DSM: Diagnostic Statistical Manual

SNP: single nucleotide polymorphism

AUD: Alcohol Use Disorder

AD: Alcohol Dependence

PFC: Prefrontal Cortex

NAc: Nucleus Accumbens

VTA: Ventral Tegmental Area

GWAS: Genome-Wide Association Study

ADH: Alcohol Dehydrogenase

ALDH: Aldehyde Dehydrogenase

DNA: deoxyribonucleic acid

RNA: ribonucleic acid

GWAS: Genome-Wide Association Study

GSCAN: GWAS & Sequencing Consortium of Alcohol and Nicotine use

AUDIT: Alcohol Use Disorders Identification Test

IEA: Intermittent Ethanol Access

QTL: Quantitative Trait Locus

bQTL: Behavioral Quantitative Trait Locus

eQTL: Expression Quantitative Trait Locus

LD: Linkage Disequilibrium

EW-dmGWAS: Edge-Weighted Dense Module searching for Genome-Wide Association Studies

PPI: Protein-Protein Interaction Network

LOD: Logarithm of the Odds

RI: Recombinant Inbred

BXD: C57BL/6J x DBA/2J recombinant inbred

IEA: Intermittent Ethanol Access

CEA: Chronic Ethanol Access

DO: Diversity Outbred

WGCNA: Weighted Gene Co-expression Network Analysis

List of Tables

Chapter 4

Chapter 5:

List of Figures

Chapter 4:

Chapter 5:

ABSTRACT

GENOME-WIDE SYSTEMS GENETICS OF ALCOHOL CONSUMPTION AND DEPENDENCE

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctorate of Philosophy from Virginia Commonwealth University

By

KRISTIN M. MIGNOGNA

Directed by: MICHAEL F. MILES, MD, PhD

Professor, Departments of Pharmacology & Toxicology, Neurology, and Human & Molecular Genetics

Alcohol use disorder is a debilitating disease for which widely effective treatment is not yet available, because the exact biological mechanisms that underlie this disorder are not completely understood. One way to gain a better understanding of these mechanisms is to examine the genetic frameworks that contribute to the risk for developing this disorder. However, because of its genetic complexity, genetic association studies have traditionally struggled to identify replicable affective loci. Even more recent, better-powered studies have yet

to account for the 50% heritability estimated by twin studies. Furthermore, these studies do not provide information regarding the mechanistic frameworks through which the identified loci affect the trait. Gene expression studies can both provide insight into these frameworks, but also identify groups of co-regulated and co-functioning genes that may be individually identified by different association studies. Gene expression in the brain is difficult to study in humans, due to lack of experimental control. Mouse models can be used to examine gene expression patterns in the brain, in the context of ethanol treatment and consumption, in an experimentally controlled environment. However, human behavior is more complex than mouse behavior, and the genetic drivers of behavior interact with the environment.

The study presented in Chapter 1 capitalizes on the benefits of the behavioral complexity of human samples and the experimental capabilities provided by mouse models by co-analyzing gene expression networks in the mesolimbocortical system of acute ethanol-treated recombinant inbred mice and human genetic alcohol dependence association data. This study successfully identified ethanol-responsive gene expression networks unique to the Ventral Tegmental Area, Nucleus Accumbens, and Prefrontal Cortex. Modules from each brain region were significantly overrepresented with genes suggestively associated with alcohol dependence in an independent human sample. These networks were functionally overrepresented with pathways related to actin-related activity, transcript regulation, ubiquitination, and Syndecan and Wnt signaling, implicating the involvement of these pathways in the association between initial alcohol sensitivity and dependence. These results indicated that gene expression networks in mouse models are informative for identifying mechanistic networks relevant to the risk for developing dependence, and for bridging the gap between human association studies.

The idea study population would allow for fine genetic mapping and gene expression network identification in the same organism. Outbred mice, such as the Diversity Outbred (DO) sock, are more genetically, and therefore behaviorally diverse than traditional inbred mouse strains. The genomes of these mice contain variants not present in other mouse strains, but also provide for mapping precision similar to that in humans, due to relatively small linkage disequilibrium blocks resulting from several generations of outbreeding. However, to our knowledge, ethanol drinking behaviors have not yet been genetically studied in DO mice, or the eight progenitor strains used to create this outbred stock. Our study in Chapter 3 characterizes voluntary ethanol consumption and preference under intermittent (shown to model the relapse-like progressive increase in drinking seen in the beginning stages of dependence development) and chronic ethanol access paradigms in each of the progenitor strains. This study determined that these phenotypes are highly heritable amongst these strains, and therefore likely genetically mappable in the DO mice. In Chapter 4, we identify several previously identified, and one novel, genetic loci associated with these traits under an intermittent access model, with unprecedented precision. One of these loci replicated results seen in a human association study of alcohol consumption, and provided insight to the potentially contributing genes. Finally, in Chapter 5, we identified gene expression networks in these mice whose expression levels were associated with ethanol exposure and consumption, and were overrepresented with genes associated with consumption in human association studies. We also mapped genetic loci that contributed to the gene expression levels of these networks, some of which overlapped with the behavioral loci, indicating that the functions represented by these modules mediate the relationship between the genotypes in that region and drinking behaviors. The functions related to these modules largely involved actin-mediated synaptic remodeling, neurite outgrowth, and ubiquitination. Overall, our

studies revealed neuroplastic and ubiquitin-related genes pathways involved in alcohol consumption in mice and humans, and that likely contribute to the risk for developing dependence.

Introduction

Not only can alcohol misuse cause liver disease, pancreatic disease, chronic neurological problems, and even death (3), but it can also affect the population at large, via belligerent behaviors and drunk driving. In fact, alcohol misuse is the third leading preventable cause of death in the United States (4), and the fifth leading risk factor for premature death and disability, globally (5). Alcoholism can be defined in many different ways, but the gold standard for Alcohol Use Disorder (AUD) diagnosis is the Diagnostic Statistical Manual (DSM) (6). As annotated in the DSM-V, AUD spans the spectrum of problematic alcohol use, from abuse (sometimes called misuse) to dependence, with varying levels of severity: mild, moderate, and severe, marked by exhibiting 2, 3-5, and at least 6 of the 11 possible symptoms, respectively (7). The lifetime prevalence of AUD has been estimated to be as high as 29.1% in the United States, and 13.9% for specifically severe AUD (5). The 11 possible symptoms are as follows:

1) drinking more or for longer stretches of time than intended

2) failing to reduce or prevent oneself from drinking, despite desire to do so

3) spending a substantial amount of time drinking and/or recovering from drinking

4) preoccupation with desire to drink

5) failing to perform duties associated with life roles, due to drinking

6) drinking despite knowledge that it is harming one's personal relationships

7) reducing participation in normally desirable leisure activities in order to drink

8) experiencing potentially harmful situations due to drinking, more than once

9) continuing to drink despite known psychological, neurological, and/or systemic harm

10) exhibiting tolerance (needing to drink more to feel the same positive effects)

11) exhibiting withdrawal symptoms (insomnia, nausea, tremors or seizures, restlessness, sweating, heart-racing, or hallucination) (7).

Older versions of the DSM considered abuse and dependence to be distinct disorders, with abuse reflecting more of what could be considered to be excessive use, and dependence reflecting more of what could be considered addictive use. In the DSM-IV, Alcohol Abuse was defined as the presence of at least one of AUD symptoms 5, 6, and 8, or multiple instances of arrest or legal issues related to alcohol use. DSM-IV Alcohol Dependence (AD) was defined as the presence of 3 or more of AUD symptoms 1-3 and 7-1 (7). However, the DSM-V recognizes that abusive use often gives rise to dependence by merging the diagnoses under a single umbrella that better reflects the developmental process of dependence.

Although the initial motivation behind drinking and the developmental timescale varies between individuals, the developmental pathway of dependence generally follows 3 stages: acute use, chronic use, and compulsive use (8). At first, one drinks for the positive reinforcing effects of alcohol largely mediated by the dopaminergic reward pathway, such as social inhibition, feelings of relaxation, or anxiolysis. After continuous use, the brain's neurochemistry begins to change, causing tolerance, leading the frequency and amount of alcohol drinking to increase (i.e. chronic use or abuse) (9). This consistent elevated level of drinking leads to not only increased changes in neurochemistry, but also eventually reroutes neuronal circuits, via synaptic plasticity (10). It is at this point at which people begin to feel withdrawal and craving symptoms during periods of abstinence from alcohol (10)These symptoms are the hallmarks of what is colloquially known as "alcoholism", "addiction", or "dependence" (6).

Despite the clear widespread personal and interpersonal disruption this disorder causes, a truly effective method of reducing its impact, particularly after onset, has yet to be discovered. Although psychological and pharmacological therapies are available for AUD, the short-term remission rate is only 43% after treatment, with the long-term relapse rate as high as 80% (11). Additionally, there are only 3 pharmacological treatments officially labeled for AUD treatment (Naltrexone, Disulfiram, and Acamprosate). Naltrexone and Disulfiram both work by disrupting the association between alcohol and positive reinforcement, by acting on two very distinct pathways. Naltrexone (a selective mu-kappa-opioid antagonist) decreases the acute dopaminergic reward response to alcohol (12). Disulfiram (an acetaldehyde dehydrogenase inhibitor) increases the systemic negative physiological response to alcohol by preventing the metabolism of acetaldehyde into acetic acid (12). The resulting accumulation of acetaldehyde in the blood results in several unpleasant symptoms such as facial flushing, nausea, dizziness, and headache (13). Despite the disruption of association between alcohol and acute positive reinforcement, the symptoms of craving and withdrawal still remain (Fig. 1). Acamprosate (an NMDA inhibitor and $GABA_A$ agonist), on the other hand, acts by decreasing the symptoms of withdrawal (12). These symptoms include, but are not limited to: irritability, anxiousness, vomiting, headache, excessive sweating, tremors, seizures, delirium tremens, and sometimes hallucination. These symptoms can begin after only hours of abstinence, and can continue for up to a week (14). However, Acamprosate is a relatively non-specific drug that can actually increase psychiatric symptoms that often drive people to drink, such as depression and anxiety. Furthermore, it has no direct effects on craving, which can occur throughout the rest of the person's life, even after withdrawal symptoms are no longer present (14). It would therefore be beneficial

to create a therapeutic drug that specifically targets biological pathways involved in alcohol craving. Unfortunately, these specific signaling cascades and neurological pathways have yet to be identified. However, we do know of several neurotransmitters and brain regions that are likely involved, particularly those that compose the “reward system”.

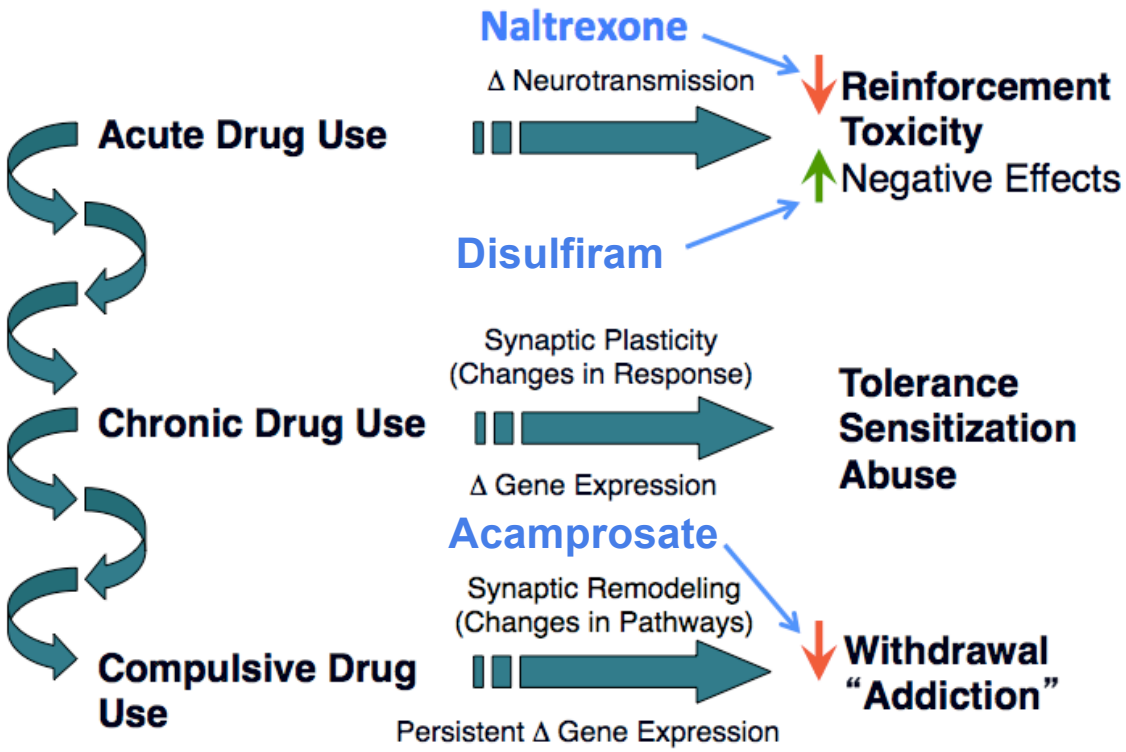


Figure 1: The three stages of alcohol dependence development (left) and the mechanisms by which they confer neurological changes and therefore risk for progressing to a later stage of development. Blue text and arrows represent approved pharmacotherapies for dependence, and in stages in which they interfere. Red and green arrows indicate the actions of the medications with respect to the alcohol-related traits and symptoms on the right.

One general neurological pathway that is known to be involved in reward and craving are the mesolimbocortical reward pathway, a dopaminergic pathway originating in the Ventral Tegmental Area (VTA), and projecting to the Nucleus Accumbens (NAc), and Prefrontal Cortex (PFC) (14). This pathway is acutely activated by alcohol, with a resulting release of dopamine in the NAc and PFC. This is thought to cause or mediate the rewarding properties of alcohol (14). However, after long-term use, these reward pathways are blunted, basally. This is due to over-stimulation of neurons, triggering negative feedback that results in the reduction of post-synaptic dopaminergic receptors (14). This means that an alcohol dependent individual's neurons responds less to dopamine than that of a non-dependent individual, which is thought to contribute to cravings, particularly in the PFC (14). The Prefrontal Cortex's relationship with craving corresponds with increased levels of dopamine, with respect to baseline, indicating that the involved pathways are similar to (or the same as) those related to reward (15). The same type of feedback response occurs in many other pathways, resulting from long-term alcohol use. These pathways involve neurotransmitters such as GABA (gamma-aminobutyric acid) and serotonin (14). The effects of other neurotransmitters, such as glutamate, are acutely inhibited by alcohol have the opposite response: the brain's response to those neurotransmitters increases after long-term use (14).

The role of the Prefrontal Cortex in drinking behaviors extends beyond initial reward. It is associated with the behavioral disinhibition and impulsivity, both of which contribute to alcohol dependent individuals' inability to control their drinking behaviors (16). It has been shown that people with greater average drinking levels and behavioral disinhibition correlate had decreased medial PFC activity while performing reward-seeking behaviors, indicating involvement of the medial PFC in controlling reward-seeking behaviors, including alcohol

consumption (17). The effects of alcohol on disinhibition and impulsivity span from the acute effects of intoxication to long-lasting effects caused by chronic drinking (14). The long-term effects play a role in the inability of individuals to control their drinking behaviors when they experience craving. It has been proposed that the long-term effects of alcohol on prefrontal cortical mediated decision-making and loss of regulation of limbic anxiety are mediated by neuroplasticity, demyelination, and neurodegeneration (14) (9) (15). In fact, alcoholic individuals with greater amounts of cortical white matter have shown the ability to maintain sobriety for longer periods of time than those with less (18). Finally, this damage to the PFC leads to disregulation of the limbic system and increased levels of glucocorticoids in the PFC, which leads to an increase in anxiety in the absence of alcohol, contributing to the pathological desire to consume alcohol for its anxiolytic effects (19). Overall, the PFC plays an important role in linking alcohol-induced dopaminergic reward with longer-term symptoms of alcoholism, including abstinence-induced anxiety, craving, behavioral disinhibition, and impulsivity.

Although a lot is known about the neurotransmitters and brain regions that are involved in alcoholism, many of the exact mechanistic pathways involved in the alcohol-induced changes that occur in neurons have yet to be discovered. Importantly, little is known about the susceptibility and sensitivity of each individual's neural circuits and cognitive, psychiatric, and behavioral responses to those changes. It is well known that this susceptibility runs in families, which could be driven by environment or genetics (20). Twin studies have explored the source of this familial clustering, and have estimated a heritability of roughly 50%, meaning half of the variance in alcohol use disorders in the population can be attributed to genetic differences and the other half to the environment (21) (22).

Genome-Wide Association Studies (GWAS), which query alleles at loci across the entire genome for association with a phenotype, have identified several genes that are significantly associated with alcohol consumption and dependence (Table 1). However, aside from a small handful of genes, few results have been significant after correction for multiple testing or replicate significance across samples (23). (Replicated genes include: *ALDH2* for dependence; and *AUTS2*, *KLB*, for alcohol consumption; and *ADH1C* and *ADH1B* for both phenotypes.) One of the leading theories for this "missing heritability" is that hundreds, or even thousands, of genes with individually small effect sizes are acting together to contribute to overall risk, and most studies to date have lacked the necessary power to detect these loci (24). As sample sizes in such studies become larger, some novel loci (such as *COL6A3*, *CADM2*, *DNAJB14*, *FAM69C*, *GCKR*, *JCAD*, *SLC39A13*, and *STPG2*) are being detected and others are being replicated (such as *ADH1B*, *ADH1C*, *AUTS2*, and *KLB*) for alcohol consumption and AUDIT (Alcohol Use Disorders Identification Test) non-problematic (AUDIT-C) and problematic (AUDIT-P) consumption scores (25) (26). However, even these studies have yet to reach the expected large numbers of significant loci. Furthermore, these studies do not implicitly provide mechanistic information regarding the relationship between the alleles and the trait.

One alternative method of exploring genetic pathways that underlie alcohol-related traits is to study gene expression patterns in relevant tissues, such as brain regions that are known to be involved in addiction, like the Nucleus Accumbens, Amygdala, Prefrontal Cortex, and Ventral Tegmental Area. One other potential reason for the general lack of success in GWAS is phenotypic admixture. AUD is a complex behavioral disorder that may be driven by several different etiologies. Individuals have many different reasons for

drinking, such as social anxiety, social conformation, self-medication for depression or anxiety, mental escape from traumatic memories or environment, or one of many other potential motivations. Additionally, different biological pathways, and therefore different genes, may be responsible for the progression from chronic consumption to dependence between people. Including all of these individuals in the same sample can mask the effects of genes relevant to dependence only in subgroups of the studied population. An additional benefit of gene network analysis is that it can reveal groups of genes that function together in some of these distinct pathways, whose constituent genes display sub-significant GWAS signals due to this phenotypic admixture.

Table 1: GWAS Summary Table

Study	Phenotype	Sample Ancestry	Discovery Sample Name	Discovery Sample Size	Genes Containing SNPs with p<5e-8	Within-Study Replication
Treutlein et al., 2009	AD case/control	German	German GWAS	1,024 cases; 996 controls	*PECR*	NA
Bierut et al., 2010	AD case/control	European American and African American	SAGE	1,235 EA and 662 AA cases; 1,433 EA and 449 AA controls	No Significant SNPs	No significant replication
Edenberg et al., 2010	AD case/control	European American	COGA	847 cases; 552 controls	No Significant SNPs	No significant replication
Kendler et al., 2011	AD factor score	European American and African American	MGS2	2,375 cases; 3,393 controls	No Significant SNPs	No significant replication
Heath et al., 2011	AD symptom factor score	Australian	OZALC	1,333 cases; 1,268 controls	No Significant SNPs	No significant replication
Schumann et al., 2011	Average g(ethanol) per day	European	Combined samples from 12 studies	26,316	*AUTS2*	Replication in dataset from independent European samples from 7 studies
Frank et al., 2012	AD case/control	German	German GWAS	1,333 cases; 2,168 controls	*ADH1C*	Replication in COGA
Zuo et al., 2012	AD case/control	European American	SAGE and COGA	1,409 cases; 1,518 controls	No Significant SNPs	No significant replication
Wang et al., 2013	DSM-IV AD symptom count	European American	COGA	2,322	No Significant SNPs	No significant replication
McGue et al., 2013	AD factor score	European American	MCTFR	7,188	No Significant SNPs	No significant replication
Park et al., 2013	AD case/control	Korean	Korean GWAS	117 cases; 279 controls	*ADH7* and *ALDH2*	Replication in independent Kroean sample
Quillen et al., 2014	AD case/control	Chinese	Chinese GWAS	102 cases; 212 controls	*ALDH2*	NA
Gelernter et al., 2014	DSM-IV AD symptom count; AD case/control	European and African American	Yale-Penn and SAGE	Yale-Penn: 2,379 EA and 3,318 AA SAGE: 2,752 EA and 1,311 AA	*LOC100507053, ADH1B,* and one intergenic SNP	Replication in independent Yale-Penn/SAGE sample and German GWAS case-control sample
Edwards et al., 2015	AUDIT + DSM-IV AD symptom factor score	British	ALSPAC	4,304	One intergenic SNP	No significant replication
Schumann et al., 2016	Average g(ethanol) per day	European	Combined samples from 20 studies	70,460	*KLB*	Replication in dataset from independent European samples from 15 studies
Adkins et al., 2017	AD case/control	Irish	IASPSAD	706 cases; 1748 controls	*COL6A3*	No significant replication
Clarke et al., 2017	Average drinks per week	British	UK Biobank	112,171	*ADH1B, ADH1C, ADH5, ARID4A* (excluding never-drinkers), *CADM2, CTNNA2* (males only) *DNAJB14, FAM69C, GCKR, KLB, STPG2*, and two intergenic SNPs	NA
Walters et al., 2018 (BioRXiv)	AD case/control	European and African American	PGC (Combined samples from 14 case/control studies and 9 family studies)	11,569 EU cases and 3,335 AA cases; 34,999 EU controls and 2,945 AA controls	*ADH1B* and one intergeneic SNP	Replication of *ADH1B* across EU-only and AA-only PGC samples
Sanchez-Roige et al., 2019	AUDIT-P and AUDIT-C scores	European	UK Biobank and 23andMe	UK Biobank: 121,604; 23andMe: 20,328	*ADH1B, ADH1C, GCKR, JCAD, KLB, SLC39A13*	NA
Kranzler et al., 2019	AUDIT-C scores and AUD (ICD abuse + dependence)	European American, African American, Latino American, East Asian American, South Asian American	Million Veteran Program	242,317 EA; 61,762 AA; 15,864 LA; 1,565 EAA; and 228 SAA	AUD only: *ADH4, SIX3, chr10q25.1*, and *DRD2*; AUD and AUDIT-C: *ADH1B, ADH1C, FTO, GCKR,* and *SLC39A8*; AUDIT-C only: *KLB, VRK2, DCLK2, ISL2, IGF2BP1, BRAP, PPR1R3B, BAHCC1*, and *RBX1*	Replication of *ADH1B* across EA-only and LA-only ancestry groups
Liu et al., 2019	Average drinks per week	Wide varietied meta-sample	GSCAN	941,280	99 significant SNPs	NA

A summary of prominent Genome-Wide Association Studies conducted between 2009 and 2019, adapted from Hart et al., 2015.

Gene expression changes are thought to play a major role in the brain's response to short- and long-term exposure to alcohol, ultimately mediating neuroplastic changes that can be permanent even after drinking cessation in long-time alcoholics (9). In fact, several studies have identified alcohol-responsive genes and gene expression networks, meaning the genes' expression levels change in response to alcohol (27) (28) (29) (30) (31) (15). Many of these networks showed overrepresentation of genes having similar functionality, indicating that those functional pathways may be associated with the brain's response to alcohol. Such results include a wide array of functions largely involving regulatory, immune, and neurological pathways, with some examples being: ribosomal function, oxidative phosphorylation, RNA processing, GTPase activity, actin cytoskeleton regulation, immune response, calcium signaling, opioid signaling, neuroplasticity, neuroinflammatory response, and regulation of neurotransmission (28) (27) (31) (29). Findings like these display the utility of genomics in gaining a better understanding of the neurological mechanisms involved in AUD.

Using this information in combination with expression data can paint a more complete picture of congenital differences associated with a higher risk for developing AUD. Information from such studies will eventually lead to a deeper understanding of the disease process, both in a generalizable way and in a patient-specific way, lending itself to the development of more effective, precise, and personalized pharmacotherapies. However, conducting gene expression analysis on human brain tissue can be difficult to conduct and interpret, due to: inability to collect brain tissue from living participants; unmeasured and uncontrollable environmental variables; differing causes of death; inability to control alcohol

access in frequency, amount, or timing with respect to death; variability in post-mortem time interval prior to tissue collection; and the small sample sizes.

Mouse models provide control over these variables and availability of larger sample sizes, while allowing for the examination of mammalian tissue collected under experimental conditions. In fact, many of the previously conducted alcohol-related brain-region-specific studies mentioned above have identified the aforementioned alcohol-responsive genes in mouse brains (27) (28) (31). However, many mouse strains and stocks are genetically limited in mapping (i.e. localization of a phenotype-contributing allele on a chromosome) resolution, due to large LD blocks across the genome caused by lack of chromosomal recombination (as a result of few cross-breeding generations) (32). They are also limited in genetic and behavioral variability, meaning their genomes do not consist of as large of an array of alleles as the human genome, due to the limited number of inbred strains initially bred (i.e. founder or progenitor strains) to form the crossed strains (32). This means that these strains are not typically ideal for precise mapping of genetic loci that contribute not only to the behavior in question, but also to variability in gene expression.

The Diversity Outbred stock breeding scheme mitigates these issues by incorporating 8 unique progenitor strains, including 3 wild-derived strains (which provide additional alleles not seen in most laboratory strains), into different breeding patterns followed by several generations of outbreeding. The genomes of the resulting mice is more similar to the human genome than that of other strains, in that it contains roughly 45 million SNPs (compared to the 10 million in the human genome) (33) and smaller LD blocks, resulting in higher behavioral variability, more potentially contributing loci to be examined, and more precise

mapping resolution (34) (35). However, to our knowledge, alcohol-related behaviors have yet to be studied in these mice.

This study aims to identify behavioral Quantitative Trait Loci voluntary ethanol consumption in the DO mice, and to examine related gene expression networks in the Prefrontal Cortex. Before this could be done, however, the phenotype needed to be characterized in each of the 8 progenitor strains, to confirm heritability and therefore genetic mappability. Specifically, we characterized voluntary ethanol consumption in these strains under Chronic and Intermittent Ethanol Access models, and determined high heritability for ethanol consumption and preference for ethanol over water. This experiment is described in *Chapter 3*. To our knowledge, this was the first study to characterize voluntary alcohol consumption in several of the 8 progenitor strains, and to estimate the heritability of an alcohol-related trait amongst them.

Results from the DO progenitor strain study indicated that ethanol consumption and preference were not significantly higher under the intermittent than chronic ethanol access paradigm for these particular strains. However, this could be due to small sample sizes, particularly for the C57BL/6J mice, in which intermittent access has previously shown to increase ethanol consumption over chronic access (36) (37). Additionally, the heritability for these traits was higher under intermittent access, indicating that it may be more genetically driven. Therefore, for the experiment described in *Chapter 4*, we exposed DO mice to an intermittent ethanol access paradigm in order to map genetic loci that contain alleles that contribute to the variation in this behavior. Because of the genetic and behavioral diversity in the DO genome, we were able to map several unprecedentedly low, replicated and novel,

genetic loci associated with ethanol consumption, preference for ethanol over water, and preference for a higher concentration of ethanol over lower concentration.

One of the benefits of using model organisms, as opposed to humans, for genetic mapping of behavioral phenotypes is that gene expression in the brain can be examined in the same samples in which the genetic associations were identified. This allows for the investigation of potential neurological pathways through which the associated loci affect the phenotype. In order to gain insight into pathways that may potentially mediate the association between alleles and their associations with voluntary drinking behaviors, we analyzed gene expression levels in the PFC of the DO mice (using RNAseq), as detailed in *Chapter 5*. Specifically, we 1) mapped gene networks in the PFC and determined which ones correlated with ethanol consumption or preference, 2) compared expression levels between the 100 highest and 100 lowest drinkers to identify alcohol-regulated or alcohol-regulating genes, and 3) tested for QTL for expression values of each gene across the genome, and identified those that regulated genes without the PFC networks. We then identified QTL for each network's first principal component (also known as the "Eigen gene", or theoretical gene whose expression levels account for the maximum possible variance in the expression levels of all other genes in that network) (38). Finally, we tested for physical overlap between the expression QTL (both network-wide and for individual genes within the network) and behavioral QTL, and tested for overrepresentation of genes contained in behavioral QTL intervals in networks. To our knowledge, this was the first to examine alcohol-associated gene expression patterns in the DO mouse brain. These results identified candidate PFC-related mechanisms related to ubiquitination and neuroplasticity through which the

significantly overrepresented or overlapping behavioral QTL may take effect on drinking behaviors.

For our final study, as opposed to utilizing a complex mouse model, we took the approach of capitalizing on the benefits of each model organism gene expression data and human genetic association data. As described in *Chapter 2*, we did so by co-analyzing human AD GWAS summary statistics with gene expression data from the Ventral Tegmental Area, Nucleus Accumbens, and Prefrontal Cortex of acute-ethanol-treated BXD mice. For this particular experiment, because allelic association statistics were gleaned from human data and no behaviors were measured in the mice, there was less of a need for the QTL mapping and behavioral benefits provided by outbred strains. We therefore took advantage of the reliability provided by technical replicates by utilizing BXD recombinant inbred mice, as opposed to DO mice. This is thought to be the first study to directly simultaneously co-analyze human GWAS, protein interaction, and mouse gene expression data to identify gene networks associated with the brains initial response to alcohol and the development of dependence. By identifying networks that were overrepresented with nominally (or suggestively) AD-associated genes from an independent yet similar GWAS dataset, this study also provided evidence that, although different genes may be identified between GWAS samples, these genes may take part in the same gene-region-specific biological pathways.

Chapter 1

Background

Human Genetic Studies of Alcohol Consumption and Dependence

The heritability of Alcohol Use Disorder has been estimated to be roughly 50%, by twin studies (22) (21), meaning that half of the variability of this disorder in the population can be attributed to genetic differences. However, it has been argued that twin studies likely overestimate heritability, due to violated assumptions, such as the "equal environment assumption" and the assumption of no genetic dominance, epistasis (gene-gene interactions), or gene-environment correlations or interactions (39) (24). The "equal environment assumption" states that monozygotic twins and dizygotic twins have an equal amount of shared environment between twins, allowing for all increased similarities between monozygotic twins (compared to dizygotic twins) to be attributed to increased genetic similarity (39). This could possibly lead to overestimation of heritability, if monozygotic twins have more similar environments between them than dizygotic twins (39). However, this has been shown to only have a mild effect on heritability estimates (39), and a relatively recent meta-analysis of several twin and adoption studies estimated 49% heritability for alcohol use disorders (21). Furthermore, SNP-heritability, which is estimated based on measured relatedness according to SNP similarity, of AD has been reported to be as high as 33% (40). Given that SNP-heritability does not account for complex epistasis (interactions between genes) beyond additive SNP effects, or other types of genetic similarities such as epigenetic (regulatory chemical groups bound to DNA) and gene expression similarities, SNP-heritability may be an underestimation. However, it likely estimates the majority of the variance that can be captured by GWAS.

Many of the first attempts to identify specific genes associated with the disorder are known as linkage studies (41) (42) (43) (44). The linkage approach examines the co-inheritance of a phenotype with chromosomal segments within families (45). If a specific segment is consistently inherited with the phenotypes of interest, then the effective locus is thought to be in linkage disequilibrium (LD) (or co-inherited more often than chance, due to physical proximity preventing frequent genetic recombination) with the target locus that is used to identify that chromosomal region (45). Some of these studies tested loci across the genome, and others tested a specific gene of interest (41) (42) (43) (44). While informative, these studies had very limited genomic resolution, and results often did not replicate across studies or did not explain a significant portion of the population variance (46).

Another common method, known as the "candidate gene study", was to test alleles within a single candidate gene for association with a phenotype (i.e. alcoholism) in a population of unrelated individuals. These genes were chosen in a hypothesis-driven manner, based on previously established associations or alcohol-related metabolic or neurological pathways (ex. dopamine-, GABA-, or serotonin-related) (47). Results often did not replicated across candidate gene studies and linkage studies, or even across multiple candidate gene studies (47) (41). Consistent results were found for alcohol and aldehyde dehydrogenase genes, which code for proteins that are directly involved in the metabolism of alcohol into acetaldehyde, and acetaldehyde into acetic acid, respectively (42) (48) (49). Any mutations that cause ADH to work more effectively, or ALDH to work less effectively, cause a buildup of acetaldehyde in the body. Just as with ALDH-inhibiting AUD medication Disulfiram, these types of genetic variation result in a plethora of negatively reinforcing side effects due to the increase in aldehyde dehydrogenase. Therefore, such alleles are actually protective

against the development of AUD, meaning individuals who carry these mutations are less likely to develop AUD because they are less likely to drink as much as someone who does not experience those negative side effects at such an extreme level (13). *ALDH2*, in particular, has been identified primarily in samples of Asian ancestry, because the loss-of-function mutation is the major allele in Asian populations, and it is protective against the development of AD. This gene is a good example of how behavioral phenotypes, such as AD, challenge the conventional ideology of loss-of-function mutations being phenotypically deleterious. This scenario also suggests that AD may be the body's "wild type" response to chronic alcohol consumption, and molecularly deleterious alleles (such as loss-of-function mutations) prevent this "normal" progression from chronic consumption to dependence.

Findings regarding association between GABA receptor genes GABRA2 and alcoholism were also replicated in linkage studies (50) (42) (51). However, variation in these alleles does not account for a large proportion of the overall heritability (24) (52). Furthermore, very few genes displayed this level of consistency across studies. Over time, the failure of replication between studies, the inability to account for the heritability estimated by twin and adoption studies, and the rising theory that many genes with small effects were collectively contributing to these traits, all indicated that many of the significant findings in candidate gene studies (which assume large effect sizes for each association) were likely false positives (53) (54) (55). With this revelation came the realization that the genes we think might be responsible for these traits may not be the only ones contributing significantly to the heritability of these disorders. It was clear that a hypothesis-generating approach would behoove genetic association studies.

The theory-unbiased, hypothesis-generating genome-wide association study (GWAS), which tests for association between alleles across the whole genome and the trait of interest, is now a common method currently used to identify SNPs (single nucleotide polymorphisms) associated with psychiatric (and other complex) traits. Because these studies involve a large number of association tests, a multiple-testing-corrected *p*-value of $<5*10^{-8}$ is the generally accepted significance threshold (56). This represents a Bonferroni threshold (α/N) divided by the number of independent tests, which has been estimated to be 1.2 million, based on linkage disequilibrium between common polymorphisms (57) (58) (56). When GWASs on AD and alcohol consumption first began, and no loci were significant under this stringent threshold, it became evident that the usual sample sizes did not provide enough power (59). This supported the theory that each individual locus likely has an extremely small effect size, and only in combination with hundreds (or even thousands) of other loci do these effects amount to a significant portion of the overall heritability. Other explanations could account for the lack of findings, including the possible overestimation of heritability by twin studies. However, even granting a heritability estimate as low as 33% (as estimated by SNP-heritability) (40), as opposed to 50% (as estimated by twin studies) (22) (21), GWAS findings have yet to account for this amount of population variance. This failure to account for the expected heritability with the combined effects of individual genetic variants is known as the "missing heritability" problem (24). It is believed that missing heritability is largely due to lack of power of GWAS (also partly to epistatic (or gene-gene) interactions and the exclusion of rare variants in GWAS (because of the technical uncertainty associated with microarray genotyping methods)), as opposed to over-estimation of heritability by other methods (24) (52).

Once sample sizes reached >1,000 participants, some GWAS were able to identify SNPs significantly associated with AD (23). However, most of the associations with dependence in work prior to 2016 failed to replicate across studies, even when considering alternate SNPs that affect the same genes. The exceptions to this were the ADH (alcohol dehydrogenase) and ALDH (aldehyde dehydrogenase) genes, as seen in the early linkage and candidate gene studies mentioned above (Table 1). While ADH and ALDH associations with alcohol dependence are well-accepted and have a straightforward biological explanation, as mentioned above, they do not account for a large proportion of the population variance (24) (52).

This lack of power in GWAS of psychiatric disorders is thought to stem from the individually small effects each of the anticipated hundreds, or even thousands, of associated genes have on the trait, which collectively contribute to the overall heritability, as mentioned above (55). However, newer investigations have tackled this problem by increasing sample sizes (in order to increase power) to hundreds of thousands of individuals. The first of such studies to be published was performed by Schumann et al. (60) on a sample of ~105k participants of European ancestry. They first identified novel nominally significant ($p<1*10^{-6}$) associations with weekly alcohol consumption in 5 different genes (AUTS2, TRAF, KLB, GCKR, ASB3) in ~70k samples, 1 of which replicated in the remaining ~35k samples: KLB (β-klotho).

Another example of GWASs with large samples sizes is Clarke et al. (26), on a sample of ~112k individuals from a ~500k-person population sample of adults from the United Kingdom, called the UK Biobank. This study not only replicated previous GWAS findings for weekly alcohol consumption, including ADH genes, *KLB*, and *GCKR* (which

was of nominal significance in Schumann et al.'s study) (60), it also identified novel associations in two genes: *CADM2*, and *FAM69C*. Most recently, the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) work group combined data from multiple datasets in order to obtain a sample size of ~940k. This study identified 111 loci associated with weekly alcohol consumption, 11 of which fell within an ADH gene (*ADH1B*), as well as loci that fell within genes associated with dopaminergic and glutamatergic pathways (61).

Most of these larger studies have focused on alcohol consumption, as opposed to AD or AUD, because it is difficult to obtain deep phenotypic data from such a large number of participants, largely due to required participant time commitment and limited monetary resources (Table 1). Somewhat surprisingly, these studies may not be capturing the same biological processes as those that are involved in alcohol use disorders. A recent investigation of AUDIT (Alcohol Use Disorders Identification Test) provided evidence that alcohol consumption in the general population is not driven by the same genetic factors as problematic alcohol use (25). Rather, the genetic correlation (the similarity in contributing loci) between problem use and other psychiatric disorders, such as Schizophrenia, was higher that that between AD and alcohol consumption. This genetic correlation could be due to a causal phenotypic relationship, such as drinking to self-medicate psychiatric symptoms or drinking triggering psychiatric symptoms in those who are genetically predisposed to developing them. This relationship could also be due to shared etiology between AUD and other psychiatric disorders. Shared etiology could indicate that symptoms across diagnostic categories could be different presentations of the same disorder, or that the same biological mechanisms contribute to the susceptibility for developing AUD and other psychiatric disorders. However, this is not to say that alcohol consumption is totally genetically

unrelated to AD. The genes identified for non-problematic drinking in this study may have been tapping into loci specifically associated with social drinking, and therefore may have been reflecting traits such as social group association, social class, extroversion, and gregariousness. Therefore, we believe it is still beneficial to our knowledge of AUD to study drinking behaviors, but with refined methodology that helps preclude social drinking situations that may be less closely related to dependence.

Despite the recent progress in GWAS, it is believed that there are still many more unidentified contributing genetic variants (59) (52). Additionally, with genotypic data alone, it can be difficult to identify the biological pathways through which these variants take effect on the trait. Functional information could reveal whether or not many different variants are being identified from the same mechanistic pathways within studies, or even across samples. General functional overrepresentation of identified SNPs may not tell the whole story, as there may be groups of genes that work together in an alcohol-specific manner. Furthermore, proteins or RNA molecules that participate in these frameworks may be coded by genes that do not necessarily contain associated alleles, but could be more suitable targets for pharmacotherapy than those that have been identified by GWAS. While association studies are necessary for gaining understanding of genetic contributions to AUD, they are not sufficient. Model organisms provide a platform for studying mechanistic frameworks at a more holistic level, with all of the benefits of experimental control.

Mouse Genetics of Voluntary Alcohol Consumption

Mouse Genetic Models

Classic inbred mouse strains are all genetically identical within each strain, which is beneficial for non-genetic studies, as it provides behavioral and biological consistency. They are also useful for studies aimed at studying single genes, using an inbred genome as a backdrop to a mutation or genetic abnormality, producing what are called congenic mice (62) (63). However, these mice are not suitable for the genome-wide mapping approaches necessary for studying complex traits (such as alcohol-related behaviors), because behavioral differences between strains could be cause by any one of the millions of variants that exist between the two inbred genomes. Conversely, mice that stem from crossbreeding between these strains possess a combination of variants from the progenitor (or founder) strains, and mice bred in different lines possess different combinations. Therefore, researchers can obtain populations of genetically distinct mice, allowing for the detection of the contribution of genetic differences to a measured phenotype. If the measured phenotype is consistently similar between mice that share an allele at a specific (i.e. if mice with allele 1 have a different phenotype than mice with allele 2), despite being genetically different at many other loci across the genome, this locus likely contains one of the genes contributing to the phenotypic differences.

Another approach that researchers take to genetically manipulate mouse lines is to "outcross" two genetically distinct inbred progenitor strains, in order to obtain offspring that carry alleles from each of the two strains (F1) (64). Breeding F1 mice to each other results in F2 progeny, whose genomes contain an admixture of alleles from each of the progenitor strains, some of which they carry homozygously and others heterozygously (64). F1 mice can also be "backcrossed" to one of the progenitor to produce offspring. Offspring from such lines can be iteratively backcrossed over several generations in order to obtain mice that have

a genetic background primarily similar to the progenitor strain of choice, but possess a desired (natural or genetically engineered) allele from the other F0 (original generation) progenitor strain (64). This method requires genetic testing in each generation, so that heterozygosity of the target allele can be preserved (64). Alternatively, because of their genetic variability, F2 and backcrossed mice can be used to map quantitative trait loci (QTL), or regions in the genome that contain polymorphisms that are associated with a measured phenotype (64, 65). However, the precision with which the contributing locus can be identified is determined by the number of detectable chromosomal recombinations that have occurred in that region (66). By comparing several crossed lines, given that recombination event locations are not the same across all mice, the LD block size that can be analyzed is effectively smaller, even if the number of recombination events in each line of mice, is roughly the same.

One method of obtaining populations of mice on which this type of mapping can be performed is to inbreed intercrossed mice, via brother-sister mating, to form homozygous "recombinant inbred" (RI) lines (62, 64). For these strains of mice, two or more inbred progenitor strains are crossed to at least eh F2 generation, and then undergo a minimum of 20 generations of inbreeding, until all mice within that breeding line are genetically identical (62, 64). Different lines are created in separate breeding panels, in order to create strains of mice that are genetically identical within lines, but different between lines. These mice present an admixture of genetically moderated behaviors that can be attributed to specific loci across the genome, and are therefore useful for QTL mapping, with the added benefit of the availability of technical replicates within each line (62, 67-71). However, the limited number of crossbreeding generations prior to inbreeding still limits the mapping resolution of

many RI strains (66). Another way to increase the resolution is to examine a strain whose genome contains several recombination events compared to F1 mice, rendering the LD blocks much smaller. One example of such mice is the BXD RI strain. This strain involves the crossing of the inbred lines C57BL/6J and DBA/2J, between which there are approximately 4.8 million SNPs (72). Not only are these two strains genetically different, but C57BL/6J mice prefer alcohol at a much higher level than DBA/2J mice (62). Therefore, these lines have been used for genetically mapping alcohol-related behaviors and physiological responses (62, 69-71).

A more complex recombinant inbred strain used for behavioral genetic studies is the Collaborative Cross (CC) strain, which was created from 8 progenitor strains: C57BL/6J, A/J, 129S1/SvImJ, NOD, NZO, WSB/EiJ, CAST, and PWK (35, 73). The latter three strains are wild-derived, meaning the initial mice used to create these strains were found in the wild and never domesticated, whereas lab-derived strains were bred for behaviors that were amicable for handling. This means that wild-derived strains contain alleles that are not seen in lab-derived strains, which contribute to behaviors not exhibited by lab-derived strains (35). Because of the large number of progenitor strains and the wild-contributed alleles, mice derived from these progenitors exhibit a wide array of genetically-derived behaviors and pose the potential for the identification of behavior-associated loci that cannot be identified in most other crosses or RI strains (35). The CC mice allow for mapping of behavioral QTLs with support intervals ranging from 0.5Mbp to 15.5Mbp in length (74).

Unlike RI strains, outbred mice are each genetically distinct from the rest of the stock, much like a human population. Each mouse possesses a unique combination of haplotypes from each of the originating progenitor strains, across each chromosome. Exact experimental

replication cannot be performed in this type of mouse, because, unlike RI strains, there are no genetically identical replicates. The lack of technical replicates also decreases power, so that very large (in the hundreds) sample sizes are needed to genetically map behavioral traits. However, the unique behavioral and genetic diversity of these mice allow for mapping novel behavioral QTL with greater precision than most other strains, which continues to increase with each generation of outbreeding (35, 75). One example of this type of mouse is the Diversity Outbred (DO) stock, which originated from interbreeding 144 different CC lines (76). Several additional generations of random outbreeding (i.e. breeding of mice in pairs that are less related to each other than first cousins) led to the richly phenotypically and genetically diverse DO stock, which has already afforded the identification of behavioral trait-related loci with incredible precision (regions < 1Mbp at only the 4^{th} and 5^{th} generation of outbreeding) (35). However, to our knowledge, research on alcohol-related traits in these mice has not yet been published. Before this can be done, it is important to determine heritability and feasibility by studying the progenitor strains. However, voluntary alcohol consumption behaviors have not yet been characterized in all of the progenitor strains.

Voluntary Ethanol Consumption in Mice

As seen in the previous section, some tissue-specific gene expression studies have been conducted on human alcoholic postmortem brain tissue. However, studies such as these are difficult to conduct and interpret, due to the large amounts of environmental variation between human subjects. Such environmental variables include, but are not limited to: environmental drivers of alcohol consumption, access to alcohol, bias in self-reporting of alcohol consumption and other behaviors, timing of last drink with respect to death, and time between death and tissue harvest. Furthermore, it is difficult and time consuming to obtain

large sample sizes, due to the need for consent to use postmortem tissue for scientific research, and the inability to collect brain tissue from living participants. Unlike human studies on alcohol consumption or dependence, these environmental factors that may obscure genetic signals can be controlled in animal model genetic experiments. Although such models do not account for all of the nuances of human behavior, they are useful for identifying broadly applicable biological mechanisms underlying specific behaviors associated with a disorder.

The most widely utilized animal model for genetic studies on alcohol behaviors including consumption, is the mouse. The mouse genome was sequenced earlier than that of the rat and genetic strains and techniques for manipulating the genome are better characterized in mice. Many paradigms exist for studying alcohol-related behaviors in mice, modeling different addiction-related endophenotypes, such as acute functional tolerance, withdrawal symptoms, alcohol sensitivity of locomotor activity and coordination, alcohol-induced anxiolysis, and behavioral disinhibition (via operant conditioning) (77). However, voluntary ethanol consumption paradigms capture a more complex behavior which models alcohol consumption in humans. In such paradigms, mice have access to both water and alcohol, giving them the choice to drink either. The underlying premise is that mice learn, by associating smell and taste (or bottle placement) with physiological response, that alcohol provides feelings of reward (provided by the subsequent increase in dopamine) and anxiolysis, whereas water does not. Some mouse strains will voluntarily drink higher quantities of alcohol, and even prefer it over water, such as C57BL/6J, whereas other strains rarely choose to consume ethanol, such as DBA/2J (62, 78, 79). This indicates that the choice

to consume alcohol in high quantities, or to prefer it over water, is genetically driven in mice, just as it is in humans.

There are many variations of paradigms that produced increased voluntary ethanol consumption in mouse or rat models. Some of these include addition of taste additives, such as sucrose, while others involve prior forced exposure to ethanol, which tends to lead to higher levels of voluntary consumption. A popular version of this type of protocol is the Chronic Intermittent Ethanol Paradigm, which exposes mice to intermittent periods of ethanol vapor before voluntary access to ethanol begins (80). The downfall of such studies is that the forced exposure does not model the situations that occur in humans' lives. Some researchers have chosen to omit forced exposure, and have shown that certain strains of alcohol-naïve mice will still voluntarily consumed ethanol, after long-term access (36, 37, 79). This benefit of this type of protocol is that it models a situation more similar to the human experience with alcohol, which is most often voluntarily consumed. One downfall of this type of model is that alcohol-naïve mice often typically do not voluntarily consume as much alcohol as alcohol-pretreated mice (80).

Intermittent Ethanol Access Paradigm

Some studies have been able to achieve high levels of voluntary consumption without prior ethanol treatment, however, by imposing periods of forced abstinence (i.e. access to water only). These abstinence periods cause increases in drinking upon reintroduction, which is thought to be due to craving (81). This effect is thought to be driven by craving or withdrawal, neither of which have the chance to occur when alcohol is continuously available. In particular, intermittent ethanol access paradigms, comprised of regularly alternating periods of free-choice ethanol access and forced abstinence, have produced higher

levels of ethanol consumption than chronic (or continuous) ethanol access, despite varying durations of access and forced abstinence between studies (36, 37, 77, 79). Some strains increase their consumption over time under this type of paradigm more so than they do when provided with continuous alcohol access. However, studies have also shown that this response to IEA varies by strain, suggesting genetic involvement in consumption behaviors under intermittent access, just as it is in studies with continuous access or prior ethanol exposure (77-79).

Results from one recent study have indicated that problematic alcohol use is not as highly genetically correlated with alcohol consumption, as it is with other psychiatric disorders in human GWAS samples (25). This study used the polygenic risk score method of calculating genetic overlap, in which the top-scoring SNPs from the GWAS of one phenotype are weighted by their effect sizes and are then used to determine how well they predict the values of the other phenotype (20). However, twin studies have shown that drinking levels do genetically correlate with AD (82). It is possible that the causative SNPs that are common between consumption and dependence are not those that explain the maximum amount of variance for either of the phenotypes alone, causing twin studies and polygenic analyses to have contrasting results. Additionally, IEA paradigms are frequently used to model the progressive increase in drinking levels seen in humans during AUD development, as people are gradually exposed to drinking during intermittent periods of access, such as social gatherings (77). Mice that exhibit an ethanol deprivation-like effect are therefore seen as mice that model genetic predisposition for developing dependence. Furthermore, social drinking is not a factor in these experiments, as mice are often individually housed to prevent social interactions, and mice (quite obviously) do not exhibit

the same types of social norms as humans, in the context of drinking. We therefore believe that ethanol consumption in the early stages of this paradigm models non-problematic drinking, whereas post-long-term-exposure consumption models problematic drinking that leads to dependence. Based on this assumption, we predict that initial consumption will be driven by different genetic loci than post-exposure consumption, and we therefore test the two time periods separately, reflecting these two distinct behaviors.

Behavioral Quantitative Trait Loci for Voluntary Ethanol Consumption

In order to test exactly which genes are associated with voluntary ethanol consumption in mice, multiple studies have aimed to map behavioral quantitative trait loci (bQTLs) for ethanol consumption and preference for alcohol over water. bQTLs are regions of the genome that contain SNPs, or other types of polymorphisms, associated with the phenotype of interest. QTL analysis is similar to GWAS in theory, but is mathematically different. For GWAS, *p*-values are used to determine the significance of an association of an individual SNP with a trait. In QTL analysis, or linkage analysis, LOD (logarithm of the odds) scores are used to indicate the likelihood that a given region of DNA is associated with the phenotype. Technically speaking, it represents the $\log_{10}$ of the probability ratio that the tested DNA locus is in LD with a locus that affects the phenotype (83).

The size of a bQTL is determined by a support interval, which is akin to a physical confidence interval on a chromosome. Specifically, a 95% Bayesian support interval indicates the region of DNA in which there is 95% confidence that the effective polymorphism is present (84). Other methods, such as 1.5 LOD-drop in either direction (proximal and distal) from the top-scoring location in that region, do not specify an exact level of confidence, but rely on the shape of the LOD score peak across the chromosome.

Regardless of support interval calculation methodology, the precision of this interval is affected by LD block length and genetic diversity (i.e. the number of testable alleles across that region of DNA), such that smaller LD blocks and a larger number of alleles result in more precise (i.e. smaller) intervals. The goal is to winnow this interval down until the effective polymorphism(s) has (4) been identified.

Many studies have identified bQTL for voluntary ethanol consumption and preference in various strain crosses (85) (86) (87) (88) (89). These researchers have crossed at least one strain that voluntarily drinks alcohol at high levels (ex. C57BL6/J) with at least one other strain that does not (ex. DBA/2J), with the goal of identifying the specific regions of DNA that likely contribute to this behavioral difference. One such study, on ethanol consumption under a Drinking in the Dark paradigm, a binge-like model, in the BXD recombinant inbred strain (originating from C57BL6/J and DBA/2J) was able to achieve support intervals as small as 6Mbp (89). However, many of the support intervals remained ≥9Mbp, due to the small number of progenitor strains (and limited variation between those two strains, roughly 4.8 million), and the limited number of cross-breeding generations prior to inbreeding for the creation of inbred lines. Furthermore, to our knowledge no published studies on daily ethanol consumption or preference have achieved such precise mapping.

Some studies have narrowed down support intervals to a small list of candidate genes, for other alcohol-related behaviors, following genetic mapping in recombinant inbred lines by selectively breeding these lines for subsequent generations of "Advanced Recombinant Inbred" mice. However, these experiments require several generations of breeding in order to achieve enough recombination within that support interval to winnow the region down to an interpretable number of candidate genes. One example of this is a study performed by

Putman et al. (69), who fine-mapped a QTL for alcohol-induced anxiolysis (or reduction in anxiety-like behaviors) down to 3.5Mbp by first mapping the region in earlier BXD generations, then later taking advantage of later generations of the intercrossed strain, specifically those in which the target region had undergone recombination. However, to our knowledge, this has yet to be done for loci identified for alcohol consumption or preference. One way to avoid having to go through the multi-generational process of traditional fine-mapping in order to achieve reasonably small support intervals is to use outbred mice for QTL analysis.

Therefore, the project described in the *Chapter 3* aims to assess voluntary ethanol consumption and preference in the DO progenitor strains, prior to genetically mapping this trait in a DO population. This study had the additional advantage determining feasibility of using each of the eight progenitor strains in future voluntary ethanol consumption studies. This is particularly important for the wild-derived strains, which are being used more frequently to study complex behaviors not seen in lab-derived strains, due to their phenotypic and genetic uniqueness (32, 35, 90). To our knowledge, this study was the first to characterize voluntary ethanol intake and preference in several of these strains (specifically NZO/HiLtJ, CAST/EiJ, WSB/EiJ, and PWK/PhJ).

Genome-Wide Systems Genetics of Alcohol Use Disorder

Genome-Wide Gene Expression Analysis for Studies on Alcohol Behaviors

Systems genetics, or the study of various levels of genetics from DNA to proteins, specifically related to alcohol consumption and dependence can provide a more complete explanation of potentially pharmacologically targetable biological mechanisms that

contribute to the development or persistence of AUD. Some systems genetics studies focus on a single gene or group of genes, which is helpful for identifying specific mechanisms through which those genes affect the trait. However, for complex traits such as AUD and alcohol consumption, not all of the contributing genes are know. There is therefore the need for a genome-wide, hypothesis-generating approach to identify novel groups of genes that function together in a tissue-specific manner to affect the risk for high alcohol consumption or dependence. In fact, hundreds of genes have been identified as being differentially expressed between the postmortem brains of alcohol dependent individuals and non-dependent controls, or between brains of alcohol-exposed mice and alcohol-naïve controls (29). Whether ethanol-regulating (meaning that differences in expression cause differences in ethanol consumption or other ethanol-evoked behaviors) or ethanol-regulated (meaning that differences in ethanol exposure cause differences in gene expression), these genes likely affect the way the brain responds to alcohol, and thus the risk for becoming or remaining dependent.

Because the brain is regionally organized with respect to function and neuron composition, and therefore gene expression patterns, it is important for studies to focus on distinct brain regions to identify region-specific networks (91). The NAc, VTA, and PFC are often the focus of addiction-related genetic studies, due to their known role in the reward system (mesolimbocortical pathway) and behavioral control (specifically the PFC) (27-29, 31, 92-95). Such studies have identified regionally specific expression patterns associated with alcohol exposure or consumption, although some were common across these three brain regions (9, 27-29, 31, 93-96). Both common and the region-specific expression patterns provide information about the systematic functioning of the brain and how it interacts with

alcohol, and can help pinpoint candidate systems to target with pharmacotherapies for AUD (97).

Gene Expression Network Analysis

One method of utilizing genome-wide data for studying systems genetics, in a more comprehensive and mechanistic way, is to map gene expression networks. The underlying premise is that genes with correlated expression levels code for RNA or protein products that are likely interacting in some mechanistic way, or participating in the same biological function (98, 99). These networks are portrayed as graphical schematics, where each "node" (or circle) represents a single gene, and each "edge" (or line) represents a relationship between those two genes (Fig. 2). In this case, this relationship reflects a measurement related to expression values, such as expression correlation. These nodes and edges can be "weighted" by various gene-specific and relationship-specific parameters, respectively, or they can be binary (present or not present) with no weighting. Many different types of network algorithms exist, some of which map networks using machine learning algorithms, and others that mine existing background networks for nested sub-networks. There are also several different heuristics on which clustering of genes can rely, but here we focus on methods that cluster genes by covariance-based metrics without imposing an arbitrary association coefficient threshold.

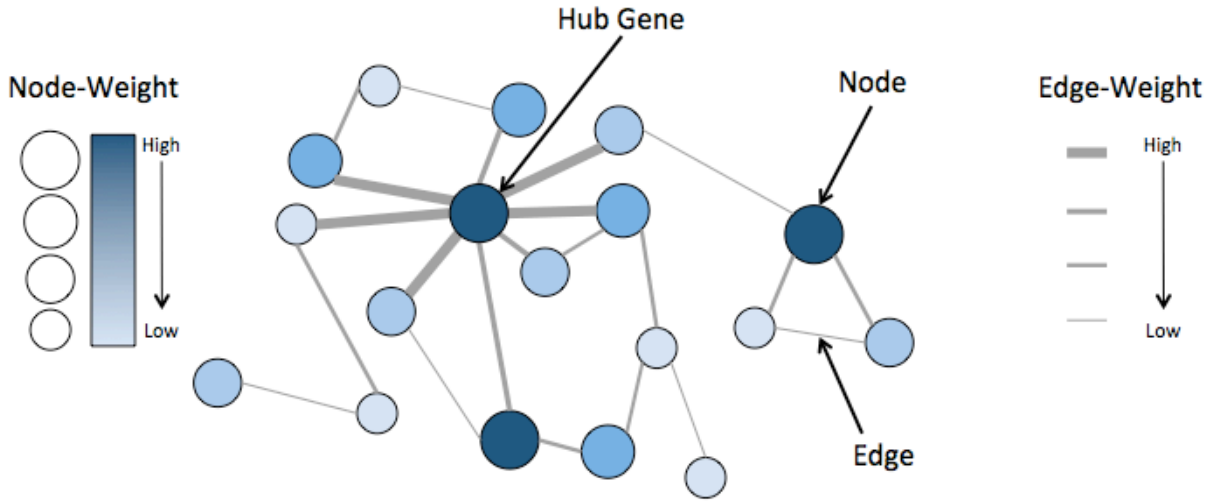


Figure 2: Nodes, or graph vertices, represent individual genes. Node-weights (often reflected by color or size) represent the effect size of the gene's allelic association with a phenotype or that node's connectivity. Connectivity metrics rely largely on the number of edges attached to a node, and sometimes incorporates the weights of those connections (or edges) into the calculation. Edges, represent biological interactions or relationships (ex. expression correlation) between genes. Edge-weights (often reflected by the thickness of the lines) represent the magnitude or effect size of that relationship. Hub genes (in blue) are nodes characterized by high connectivity.

One common method of network mapping is to use hierarchical clustering. In its most basic form, this method "clusters" genes by expression level correlation values, starting with the most highly correlated pair of genes and eventually working its way down to the smallest values in a correlation matrix of all gene pairs. Weighted Gene Co-Expression Network Association (WGCNA) is a program that hierarchically clusters genes based on Topological Overlap. For each pair of genes, the Topological Overlap value incorporates information from adjacency (i.e. transformed correlation values, or the theoretical inverse of graphical distance) between the two genes, but also their adjacency with all other genes that could potentially indirectly connect them in a network (38). The adjacency is raised to a user-defined power to create a "scale-free" network, which is a network for which the probability of a gene having *k* connections depends on a power distribution of *k* (100). Such networks are characterized by the presence of hub genes (or genes that are very highly connected to other genes, compared to the mean number of connections), making the network relatively robust in its structure, with respect to the removal of single connections (101, 102). This structure is thought to mimic naturally-occurring networks, such as biological interactions, hence the frequent use of WGCNA to map gene expression networks (28, 29, 100). In fact, this method had been used to map gene networks in brain tissue in both humans and mice, and has identified functionally enriched networks that are significantly overrepresented with alcohol-regulated genes (29, 30). Such studies provide important information with respect to candidate genes and functional pathways to be tested further for association with dependence, and to potentially eventually be targeted with therapeutic drugs to treat addiction.

An example of a network-mining algorithm used in gene expression studies is the Edge-Weighted dense module-searching algorithm for Genome-Wide Association Studies (EW-dmGWAS). This program queries a background network of binary connections (meaning two genes are either connected, or they are not, with no quantitative weighting) using independently derived gene expression-related values as weights to determine sub-network membership. This method is unique in that it is also capable of incorporating weights unique to each gene, typically a value that represents that gene's association with a phenotype, which will be discussed in more detail below. For each pair of genes that is connected in the initial network, the algorithm calculates the difference in their correlation between experimental groups. The more the experimental treatment perturbs the interaction (i.e. the larger the difference in correlation values between groups), the larger the assigned weight. Genes are iteratively added to a sub-network, following the connections of the initial network outward, based on a combined metric that incorporates its weight with each of the genes already placed in the sub-newtork (and its gene-specific phenotype-related weight, if that value is being incorporated). This algorithm is beneficial for combining data from multiple sources, as opposed to relying solely on gene expression to map networks. Its intended use is for the integration of the following three types of data, into one streamlined analysis: protein interaction data, as the initial background network; gene expression data, for the connected weights and determination of sub-network belongingness; and GWAS data, for the phenotype association-related gene-specific weights, which can also be incorporated into sub-network belongingness determination. More details on this exact implementation are described below. To our knowledge, this method has been used to map networks of genes of protein-protein interaction background network that have Alcohol Dependence GWAS p-

values (used as the gene-specific weights), but alcohol treatment-related expression values have never been incorporated into this type of analysis (103). However, previous studies have integrated alcohol-regulated gene expression and genetic association data using different method, in order to identify gene expression networks whose constituent genes contained SNPs associated with Alcohol Dependence in GWAS.

Regardless of the network-building tools used, gene expression networks identified in human studies are informative, but the contributions to differences between alcoholics and controls, and therefore their potential roles in dependence development, are ambiguous. Because of the environmental control in mouse models, and the experimental control over alcohol-related behaviors, it is possible to identify networks specifically related to endophenotype-like traits at pre-specified stages of life and dependence development. There are over 17,000 orthologs (104) between the human and mouse genome, and evidence shows that network structure is highly conserved between mice and humans (105). Several studies have identified mouse brain gene networks that respond to different levels of ethanol exposure. For example, in addition to identifying ethanol-regulated genes, Wolen et al. (27) and Smith et al. (28) identified gene networks in the PFC, NAc, and VTA of recombinant inbred mice that were responsive to acute ethanol administration (via intraperitoneal injection) and chronic involuntary ethanol vapor exposure, respectively. Two additional studies have measured gene expression differences related to ethanol consumption specifically under an IEA paradigm (the paradigm used in the studies discussed in *Chapters 3-5* of this manuscript) in C57BL/6 mice: Wolstenholme et al. (31) identified several genes with brain region-specific ethanol consumption-correlated expression levels in PFC, NAc and VTA; and Ostendorff-Kahanek et al. (95) identified 702 significantly differentially

expressed genes between IEA-exposed and control PFC. However, to our knowledge, no studies have directly integrated mouse gene expression data with human GWAS, in order to capitalize on the advantages of each.

Integration of Gene Expression Networks with Genotypic Data

A common integration method for gene expression and GWAS data is known as enrichment or overrepresentation analysis. These studies identify groups of co-expressed genes, based on gene expression data, and determine how well each network represents genes with lower (i.e. more significant) GWAS *p*-values for AD or alcohol-related behavior. For overrepresentation, a threshold GWAS p-value is chosen, the number of genes whose p-values fall below that threshold are counted, and the probability that a group of that size would contain that number or more of GWAS-suggestive or –significant genes by chance (typically via permutation), which becomes the empirical overrepresentation *p*-value. Enrichment is similarly calculated, but instead of counting the number of genes that pass some threshold, a metric (such as the mean or Fisher's combined statistic) is calculated for each group of genes, and that value is compared to values that would occur by chance. One example of a study that utilized this type of method was conducted by Farris et al., (92) gene region-specific gene co-expression networks in human postmortem PFC tissue, which contained significantly different network structures between alcohol-dependent individuals and controls. Many of these networks were also correlated with lifetime alcohol consumption in dependent participants. The most highly alcohol consumption-correlated networks were overrepresented with genes containing SNPs associated with AD in each of two GWAS datasets, meaning they contained more genes with associated SNPs than expected for a random group of the same number of genes. Similarly, Mamdani et al. (106) found that

expression quantitative trait loci (eQTL, or QTL associated with expression levels) for AD-associated gene expression networks in human NAc tissue had significant enrichment with AD diagnosis and symptom count GWAS signals. This finding is especially informative, because it indicates that some genes identified in GWAS are likely affecting AD risk via these multi-gene mechanistic networks in the brain. Additionally, many of the gene networks they identified were enriched for neuronal systems in a functional overrepresentation analysis. These results indicate that gene co-expression networks associated with alcohol consumption are informative for identifying groups of genes that collectively contribute to dependence via neurological pathways.

Other studies have integrated GWAS and human gene expression or gene network data to cross-validate behavioral genetic findings (107). For example, the Psychiatric Genomics Consortium (108) has tested for enrichment of nominally GWAS significant genes previously identified functional pathways. Their results showed shared enrichment of signals for schizophrenia, major depression disorder, and bipolar disorder in several functional categories, including histone methylation, neural signaling, and the immune system (108). Yet another approach to cross-species validation is to provide additional confirmation for GWAS significant or suggestive results for AD by showing that expression levels for such genes showed correlations with ethanol behaviors in rodent models, providing evidence of consilience of gene-trait relationships for alcohol-related behaviors across mice and humans (109). These types methods provide information regarding the function of genes that have already reached significance. However, they do not provide information about sub-threshold signals that possibly still contribute to the genetic risk and mechanisms of AUD.

Furthermore, as mentioned in *Section 1.2*, gene expression studies can be difficult to interpret when using human tissue.

Direct integrative algorithms allow for identification of gene networks coordinately weighted for GWAS significance for AD in humans and ethanol-responsiveness in model organism brain gene expression data. The EW-dmGWAS algorithm, mentioned above, can be used to directly integrate GWAS data and other biological network information in order to identify gene networks contributing to a genetic disorder, even for truly affective genes with sub-threshold significance values due to lack of power (110). One of the first studies to take this approach utilized Protein-Protein Interaction (PPI) network data to identify networks associated with AD GWAS signals. Modules mined from protein-protein interactions were scored based on node-weights based on gene-level GWAS *p*-values. This approach successfully identified AD-associated PPI networks that replicated across ethnicities. Furthermore, the resulting modules showed significant aggregate AD-association in independent GWAS datasets (103), thus demonstrating the potential utility of the method. An updated iteration of the dmGWAS algorithm, termed Edge-Weighted dense module searching for GWAS (EW-dmGWAS), allows integration of gene expression data to provide a direct co-analysis of gene expression, PPI, and GWAS data (111).

This type of approach utilizes the best of both worlds, regarding the integration of data from humans and mice. It takes advantage of the behavioral complexity in humans, while minimizing the effect of cofounding environmental factors on the biological variable, by focusing on data that explains the relationship between a trait and genotype (GWAS data), which is largely unchanging across time and environment. It also capitalizes on the ability to minimize noisy variation in time- and environment-sensitive parameters, such as gene

expression, in model organisms. For these reasons, this algorithm was used for our network studies in *Chapter 5*, for identifying alcohol-responsive, AD-associated gene networks. In order to fully capitalize on the benefits of experimental control, this analysis uses data from the ethanol-naïve recombinant inbred mice that were involuntarily exposed to ethanol via intraperitoneal injection seen in Wolen et al., (27). Although this type of experiment does not involve the measurement of an ethanol-related behavior, it minimizes error variance within groups caused by differing levels of ethanol exposure or of stress in response to behavioral testing. In contrast, the human data used for the incorporated GWASs was collected via survey, in the absence of experimental control, which creates unrealistic behavioral and environmental constraints. Finally, it has been shown that risk for developing alcohol dependence can be partially predicted by initial response to alcohol (112, 113). Whereas the IEA study reported in *Chapters 3-5* identifies genetic loci associated uniquely with naïve versus post-long-term-exposure consumption, this study allowed us to examine mechanistic pathways involved in the relationship between initial response to ethanol and its effects on long-term use.

Summary

In this Introduction, we have reviewed evidence that the mesolimbocortical pathway (including the Prefrontal Cortex, Ventral Tegmental Area, and Nucleus Accumbens) plays a role in the development of alcohol use disorder. However, many of the exact mechanisms and neural pathways involved have yet to be identified, and the current medications labeled for AUD treatment are neither precise, nor incredibly effective. Because roughly 50% of AUD's population variance can be accounted for by genetic differences, it is believed that

discovery of specific genetic associations with this disorder, or related phenotypes, can direct us towards specific biological pathways that can be targeted by pharmacotherapies. However, genome-wide association studies in humans have yet to account for as much of the trait variance as expected.

The contributing biological pathways can be examined via systems genetics approaches, which assess genetic systems from DNA to proteins. Genome-wide systems genetics approaches are particularly useful for complex traits, such as alcohol-related phenotypes, because they provide a broader, more comprehensive picture of the complex web of systems involved than single-gene or single-pathway studies. However, gene expression is tissue-specific, and it is difficult to collect gene expression data from post-mortem human brain tissue with proper environmental control or in large samples sizes. Model organisms, such as mice, provide researchers with the capability to practice experimental control and obtain large samples. In particular, the Intermittent Ethanol Access voluntary ethanol consumption mouse paradigm allows researchers to model the abstinence-driven increase in drinking seen at the beginning stages of AUD development, under an environment- and alcohol-access-controlled experimental setting. Previous genetic association studies on voluntarily alcohol-consuming mice have been limited, in both precision and number of testable loci, by the genetic simplicity of laboratory strains. These limitations likely extend to gene expression analysis, because of the effects of local and distal loci on gene expression. The Diversity Outbred stock of mice was created specifically to mitigate this issue, stemming from a complex breeding scheme involving 8 genetically distinct progenitor strains. To our knowledge, no behavioral QTL or brain region-specific gene expression studies on alcohol-related behaviors in DO mice have been published.

Therefore, this project aims to characterize voluntary ethanol consumption and dependence, and their heritability, in the 8 progenitor strains of the DO mice (*Chapter 3*). Given a high heritability (as was anticipated), indicating genetic mappability, we then sought to identify novel and unprecedentedly precise QTL for these behaviors in a sample of 600 DO mice (*Chapter 4*). We expected to see loci both unique to and common between drinking behaviors during the first and last weeks of ethanol exposure, reflecting initial and dependence-like consumption, respectively. Finally, we aimed to identify gene expression networks in the prefrontal cortex that potentially mediate the effects of these loci on drinking behaviors.

One might assume that, given the independence of the genetic influences between initial and long-term ethanol consumption in mice, that initial subjective response to ethanol would not be related to the risk for dependence. However, it has been shown that initial response to alcohol predicts risk for developing dependent use, although the exact neurological mechanisms that lead to this correlation are largely unknown (112, 113). The link between initial response to ethanol and long-term risk for dependence may be less dependent on initial drinking levels, and more dependent on the rate of acceleration of drinking levels over time, leading to differences in the likelihood of developing tolerance and the rate at which this development occurs. To gain a better understanding of this relationship, and to capitalize on the advantages of using mouse data and those of using human data, in *Chapter 2* we describe co-analysis of human Alcohol Dependence GWAS summary statistics with gene expression data from reward center-related brain regions (PFC, VTA, and NAc) acute-ethanol-treated, in the context of protein interactions. Gene expression changes in these brain regions, in response to a single dose of ethanol in ethanol-naïve mice, represents the

brain's acute response to initial alcohol exposure; whereas the human data examines the relationship between genetic loci and risk for developing dependence. The main goal was to identify gene networks associated with both initial ethanol-responsiveness and dependence.

Overall, this manuscript examines systems genetic mechanisms that underlie alcohol consumption and dependence, and the relationship between these two phenotypes. Specifically it examines networks common between ethanol responsiveness in the brain's reward pathways and alcohol dependence. It also explores the biological pathways unique to and shared between initial drinking and long-term early-dependence-like drinking . Specifically, the aims of this dissertation were to: 1) examine the relationship between initial ethanol responsiveness of gene expression to alcohol exposure and the risk for developing dependence, by capitalizing on the benefits of both model organisms and human samples (as seen in *Chapter 2*); 2) determine the genetic mappability of ethanol consumption and preference over water, by analyzing the variance in these traits between the 8 progenitor strains (as seen in *Chapter 3*); 3) to then identify novel and unprecedentedly precise bQTL that contribute to these behaviors, during naïve and post-exposure time periods, in a large sample (N≈600) of DO mice (as seen in *Chapter 4*); and 4) identify ethanol-responsive gene networks in the Prefrontal Cortex, and their respective eQTL, to gain insight to the mechanistic frameworks through which the bQTL affect drinking in DO mice (as seen in *Chapter 5*). Corresponding to these aims, we expected to find: 1) identification of novel candidate genes and gene networks involved in the relationship between initial response to alcohol and long-term risk for dependence, including genes that did not reach significance in human GWAS; 2) high heritability of ethanol consumption and preference across the 8 genetically distinct DO progenitor strains, indicating that these phenotypes are genetically

mappable in the DO mice; 3) novel and refined quantitative trait loci for ethanol consumption and preference in the DO mice; 4) gene networks in the Prefrontal Cortex that are associated with ethanol consumption levels and provide candidate biological mechanisms through which genetic variants might affect drinking behaviors.

Chapter 2

Cross-species alcohol dependence-associated gene networks: Co-analysis of mouse brain gene expression and human genome-wide association data

Introduction

Twin studies estimate that AUD is roughly 50% heritable (21, 22). Multiple rodent model studies have used selective breeding to enrich for ethanol behavioral phenotypes or have identified ethanol-related behavioral quantitative trait loci (87, 114, 115), further confirming the large genetic contribution to alcohol behaviors. Recent studies have also documented genetic factors influencing the effectiveness of existing pharmacological treatments for AD, further substantiating genetic contributions to the mechanisms and treatment of AUD (116). Genome-wide association studies (GWAS) in humans have identified several genetic variants associated with alcohol use and dependence (23, 26, 60, 117). However, they have yet to account for a large portion of the heritability estimated by twin studies. Lack of power, due to a large number of variants with small effects, is believed to the source of this "missing heritability" (24). Although recent large-scale studies have shown promise in identifying novel genetic contributions to alcohol consumption, these studies do not contain the deep phenotypic information necessary for identifying variants associated with dependence (26, 60). Further, such GWAS results still generally lack information about how detected single gene variants are mechanistically related to the disease phenotype.

Genome-wide gene expression studies are capable of improving the power of GWAS by providing information about the gene networks and biological mechanisms in which GWAS variants function (92, 107, 118, 119). Although gene expression in brain tissue has been studied in AD humans (92, 107), these studies are often difficult to conduct and interpret, due to lack of control over experimental variables and small sample sizes.

However, studies have found high conservation in gene expression correlation patterns between mice and humans, particularly in brain tissue (105). Furthermore, extensive studies in rodent models have successfully identified ethanol-associated gene expression differences and gene networks in brain tissue (28, 30, 96, 120, 121). Multiple ethanol-behavioral rodent models exist to measure different aspects of the developmental trajectory from initial exposure to compulsive consumption (77). Acute administration to naïve mice models the response of initial alcohol exposure in humans, which is an important predictor of risk for AD (112, 113). Wolen et al. used microarray analysis across a mouse genetic panel to identify expression correlation-based networks of acute ethanol-responsive genes (genes whose expression levels change after ethanol consumption or treatment), along with significantly associated expression quantitative trait loci in brain regions involved in the mesolimbocortical dopamine reward pathway -- the prefrontal cortex (PFC), nucleus accumbens (NAc), and ventral tegmental area (VTA) (120). Furthermore, specific networks also correlated with ethanol behavioral data derived from the same mouse genetic panel (BXD recombinant inbred lines) (115). Importantly, these gene expression responses to acute ethanol in BXD mice were later shown by our laboratory to have highly significant overlap with expression responses in a chronic ethanol exposure model known to mimic aspects of alcohol dependence in humans (97), and also contained a gene expression network associated with alcohol dependence that we recently identified *Gsk3b* as a potential candidate gene for treatment of alcoholism (122). Together, these results support our premise that acute ethanol-exposed rodent brain gene expression could provide insight into relevant mechanistic frameworks and pathways underlying ethanol behaviors and risk for dependence in humans.

Several studies have integrated GWAS and gene expression or gene network data to cross-validate behavioral genetic findings (107). For instance, the Psychiatric Genomics Consortium (108) tested for enrichment of nominally significant genes from human GWAS in previously identified functional pathways, and found shared functional enrichment of signals for schizophrenia, major depression disorder, and bipolar disorder in several categories. These pathways included histone methylation, neural signaling, and immune pathways (108). Mamdani et al. reversed this type of analysis by testing for significant enrichment of previously identified GWAS signals in gene networks from their study. They found that expression quantitative trait loci for AD-associated gene expression networks in human nucleus accumbens tissue had significant enrichment with AD diagnosis and symptom count GWAS signals from the Collaborative Study on the Genetics of Alcoholism dataset (107). Additional approaches have taken human GWAS significant (or suggestive) results for AD and provided additional confirmation by showing that expression levels for such genes showed correlations with ethanol behaviors in rodent models (109). Such methods are informative with respect to analyzing the function of genes that have already reached some association significance threshold. However, they do not provide information about genes not reaching such statistical thresholds, but possibly still having important contributions to the genetic risk and mechanisms of AUD

As discussed in the *Background* section, dense module searching for GWAS (dmGWAS) is an algorithm for directly integrating GWAS data and other biological network information so as to identify gene networks contributing to a genetic disorder, even if few of the individual network genes exceed genome-wide statistical association thresholds (110). The initial description of this approach utilized Protein-Protein Interaction (PPI) network

data to identify networks associated with a GWAS phenotype. Modules derived from protein-protein interactions were scored from node-weights based on gene-level GWAS *p*-values. This approach was used to identify AD-associated PPI networks that replicated across ethnicities and showed significant aggregate AD-association in independent GWAS datasets (103), thus demonstrating the potential utility of the method. A more recent iteration of the dmGWAS algorithm, termed Edge-Weighted dense module searching for GWAS (EW-dmGWAS), allows integration of gene expression data to provide a direct co-analysis of gene expression, PPI, and GWAS data (111).

Utilization of the EW-dmGWAS algorithm would allow for identification of gene networks coordinately weighted for GWAS significance for AD in humans and ethanol-responsiveness in model organism brain gene expression data. We hypothesized that such an approach could provide novel information about candidate gene networks likely contributing to the genetic risk for AUD, while also adding mechanistic information about the role of such networks in ethanol behaviors. We show here the first use of such an approach for the integration of human PPI connectivity with mouse brain expression responses to acute ethanol and human GWAS results on AD. Our design incorporated the genome-wide microarray expression dataset derived from the acute ethanol-exposed mouse brain tissue used in Wolen et al. (115, 120), human protein-protein interaction data from the Protein Interaction Network database, and AD GWAS summary statistics from the Irish Affected Sib-Pair Study of Alcohol Dependence (109). Importantly, we validated the identified ethanol-responsive and AD-associated networks by co-analysis with an additional, independent AD GWAS study on the Avon Longitudinal Study of Parents and Children dataset. Our results, although requiring further detailed investigation, could provide

important methodological and biological function insight for future studies on the mechanisms and treatment of AUD.

Materials and methods

Samples

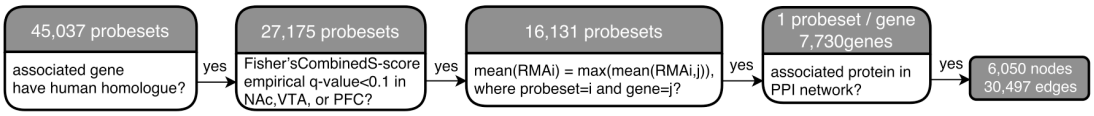
Mouse gene expression data

In order to maximize the utility of the mouse model by minimizing effects of behavioral and environmental variation, and to afford the identification of dependence-contributing genes involved in initial ethanol response, this study utilized gene expression data from ethanol-naïve mice treated with a single dose of acute ethanol. All mouse brain microarray data (Affymetrix GeneChip Mouse Genome 430 2.0) are from Wolen et al., 2012 (120) and can be downloaded from the GeneNetwork resource (www.genenetwork.org), via accession numbers GN135-137, GN154-156 and GN228-230, respectively for PFC, NAc and VTA data. Treatment and control groups each contained one sample (pooled RNA from 3 mice) from each strain and were given IP injections of saline or 1.8 g/kg of ethanol, respectively. Euthanasia and brain tissue collection took place 4 hours later. Data used for edge weighting in EW-dmGWAS analysis included Robust Multi-array Average (RMA) values, background-corrected and normalized measures of probe-wise expression, from the PFC, VTA, and NAc of male mice in 27-35 BXD recombinant inbred strains and two progenitor strains (DBA/2J and C57BL/6J).

Ethanol-responsive genes are predicted to be involved in pathways of neural adaptations that lead to dependence (120). We predicted they would also be involved in mechanistic pathways from which GWAS signals are being detected. We therefore filtered

for ethanol-responsive gene expression as done in Wolen et al.(120) prior to EW-dmGWAS so as to ensure edge weighting focused on ethanol responsiveness. Probe-level expression differences between control and ethanol-treated groups using the S-score algorithm which performs a probe-level analysis of expression between two groups (96, 123, 124) were obtained from the Wolen study (120) (S1 Table). Fisher's Combined Test determined S-score significance values for ethanol responsiveness of each probeset across the entire BXD panel, and empirical p-values were calculated by 1,000 random permutations. Finally, q-values were calculated from empirical p-values to determine the false discovery rates due to multiple testing (120). We defined an ethanol responsive gene set using a S-score probeset-level threshold of $q_{FDR}<0.1$ for differential expression, in any one of the three brain regions. Genes associated with these probesets were carried forward in our analysis (Fig 1a). Multiple probesets from single genes were reduced to single gene-wise expression levels within a particular brain region by selecting the maximum brain region-specific RMA value for each gene. After removing genes that were absent from the human datasets, 6,050 genes remained with expression values across all three brain regions (Fig 1a).

a



b

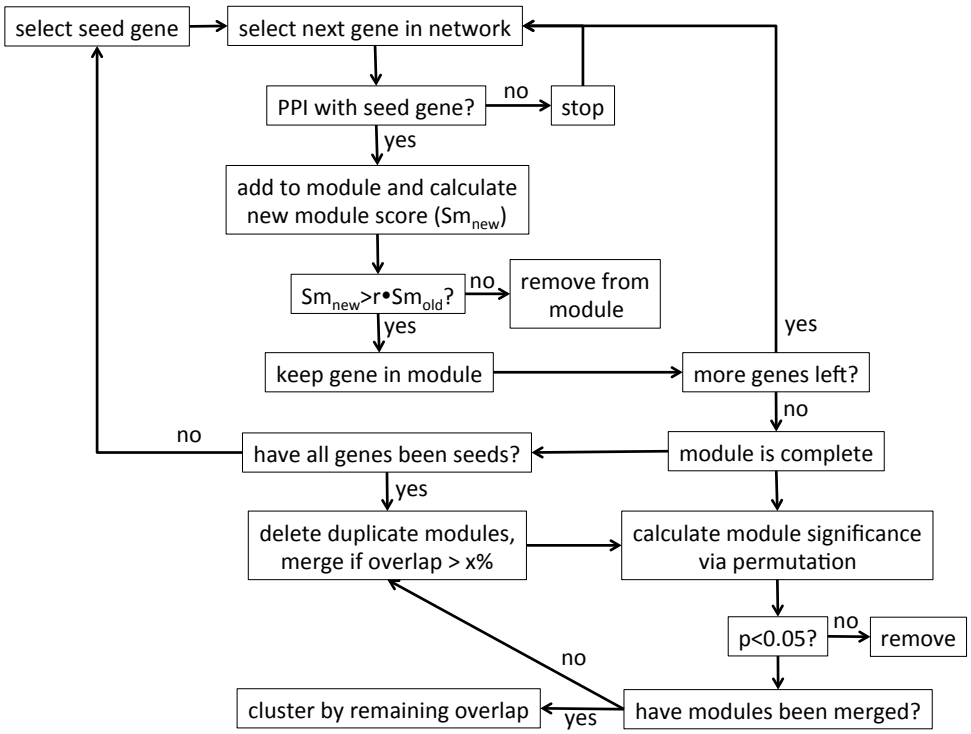


Figure 1: (a) Pipeline used to prepare the data for the present analysis. The first cell contains the starting number of genes in the BXD mouse PFC, NAc, and VTA gene expression dataset. (b) Implementations made to the standard EW- dmGWAS pipeline. Empirical p-values were calculated from standardized module scores based on a Z-distribution. Original EW-dmGWAS module score, permutation, and score standardization algorithms were used to calculate the respective Mega Modules parameters. Modules were considered to have >80% overlap if >80% of the genes in the smaller module were contained in the larger module. False Discovery Rates were calculated based on the Benjamini-Hochberg algorithm, using the "stats" package in R. Intramodular connectivity was defined as the number of edges (i.e. connections) attached to that node (i.e. gene). Eigen-Centrality was calculated using the "igraph" package in R.

Human GWAS data

Although many GWAS datasets now exist for AD, alcohol consumption and other ethanol responses, we chose two AD-related datasets for our analysis because of the phenotypic and methodological similarity between the studies and their availability at the time this work was initiated. The Irish Affected Sib-Pair Study of Alcohol Dependence (IASPSAD) AD GWAS dataset was used for the EW-dmGWAS analysis. It contains information from 1,748 unscreened controls (43.2% male) and 706 probands and affected siblings (65.7% male) from a native Irish population, after quality control (109). Samples were genotyped on Affymetrix v6.0 SNP arrays. Diagnostic criteria for AD were based on the DSM-IV, and probands were ascertained from in- and out-patient alcoholism treatment facilities. Association of each Single Nucleotide Polymorphisms (SNP) with AD diagnosis status was tested by the Modified Quasi-Likelihood Score method (125), which accounts for participant relatedness. SNPs were imputed using IMPUTE2 (126) to hg19/1000 Genomes, and gene-wise p-values were calculated using Knowledge-Based mining system for Genome-wide Genetic studies (KGG2.5) (127). This dataset was chosen because of its deep phenotyping and its theoretical consistency with findings from mouse experiments. The expression of the top-scoring genes in IASPSAD (*COL6A3*, *RYR3*, and *KLF12*) in mouse brain correlates with handling-induced convulsions, anxiety-like behavior, and acute functional tolerance to ethanol, respectively(109).

The Avon Longitudinal Study of Parents and Children (ALSPAC) GWAS gene-wise p-values were used to examine the ability of EW-dmGWAS to validate the EW-dmGWAS networks. This GWAS tested SNP association with a factor score calculated from 10 Alcohol Use Disorder Identification Test items for 4,304 (42.9% male) participants from Avon, UK.

Samples were genotyped by the Illumina HumanHap550 quad genome-wide SNP platform (128). This dataset was chosen because of its overall similarity to IASPSAD. Although the analyzed phenotypes were not identical between these two datasets, they were similar in that they both studied dependence symptoms, as opposed to non-diagnostic drinking measures. Additionally like IASPSAD, ALSPAC possessed the following important qualities: 100% of the sample was from the United Kingdom; the male to female ratio was roughly 1:1; SNPs were imputed to hg19/1000 Genomes; and gene-wise p-values were calculated by KGG2.5. No other GWAS dataset is as similar to IASPSAD to our knowledge, with respect to ancestral origin, genotyping, and phenotyping.

Protein network data

The Protein-Protein Interaction (PPI) network was obtained from the Protein Interaction Network Analysis (PINA 2.0) Platform (http://omics.bjcancer.org/pina/interactome.pina4ms.do). This platform was chosen because it includes PPI data from a wide array of databases, including: Intact, MINT, BioGRID, DIP, HPRD, and MIPS/Mpact. The *Homo sapiens* dataset was used for this analysis, due to it having much more content (166,776 binary interactions) than the mouse repository (only 13,865 binary interactions) (129, 130). Uniprot IDs were used to match protein symbols to their corresponding gene symbols (131).

Statistical methods

EW-dmGWAS

The edge-weighted dense module searching for GWAS (dmGWAS_3.0) R package was used to identify treatment-dependent edge-weighted modules (small, constituent

networks) nested within the background network(s) of non-weighted, binary interactions (https://bioinfo.uth.edu/dmGWAS/). We used the PPI framework for the background network, IASPSAD GWAS gene-wise p-values (109) for the node-weights, and RMA values from acute ethanol- and saline-exposed mouse PFC, VTA, and NAc genomic data for edge-weights (120). For the remainder of this manuscript, we will use the term "network" to refer to the background PPI framework, and "module" to refer to the resulting groups of interrelated genes nested within this larger network. By the EW-dmGWAS algorithm, higher node-weights represent lower (i.e. more significant) GWAS p-values, whereas higher edge-weights represent a greater difference in the correlation of two genes between ethanol and control groups. This is calculated by taking the difference of correlations in RMA expression values of the two genes in control vs. ethanol treated BXD lines. The module score algorithm incorporated edge- and node-weights, which were each weighted to prevent bias towards representation of nodes or edges in module score calculations. Such bias could cause some modules to be identified based almost solely on edge-weights or node-weights, as opposed to the two combined, which would defeat the purpose of integration. The respective weighting depends upon a parameter (λ) which is calculated prior to module searching, based on the entire set of node- and edge-weights and used across all module score calculations, as part of the EW-dmGWAS algorithm. Higher module scores thus represent higher edge- and node-weights. Genes were kept in a module if they increased the standardized module score (S_n) by 0.5%. S_n corresponding to a permutation-based, empirical $q_{FDR}<0.05$ were considered significant. A significant S_n (i.e. more significant q_{FDR} values) indicates that a module's constituent genes are more highly associated with AD in humans, and their interactions with

each other are more strongly perturbed by acute ethanol exposure in mice than randomly constructed modules of the same size.

Due to the redundancy of genes between modules, we modified the EW-dmGWAS output by iteratively merging significant modules that overlapped >80% until no modules had >80% overlap, for each brain region. Percent overlap represented the number of genes contained in both modules (for every possible pair) divided by the number of genes in the smaller module. We call the final resulting modules "mega-modules". Standardized mega-module scores (MM-S_n) were calculated using the algorithms employed by EW-dmGWAS. MM-S_n corresponding to q_{FDR}<0.05 were considered significant (Fig 1b). Finally, connectivity (k) and Eigen-centrality (EC) were calculated using the igraph R package for each gene in each module to identify hub genes. Nodes with EC>0.2 and in the top quartile for connectivity for a module were considered to be hub genes.

Overlap with ALSPAC

Genes with an ALSPAC GWAS gene-wise p<0.001 were considered nominally significant, and will be referred to as "ALSPAC-nominal genes". We used linear regression to test MM-S_n's prediction of mean ALSPAC GWAS gene-wise p-value of each mega-module. Given our hypothesis that EW-dmGWAS would identify alcohol-associated gene networks and prioritize them by association, we predicted that higher MM-S_n's would predict lower (i.e. more significant) mean GWAS p-values. Empirical p-values<0.017, reflecting Bonferroni correction for 3 independent tests (one per brain region): α=0.05/3, were considered to represent significant association.

Overrepresentation of ALSPAC-nominal genes within each mega-module was analyzed for those modules containing >1 such gene. For each of these mega-modules, 10,000

modules containing the same number of genes were permuted to determine significance. Empirical p-values < 0.05/n (where n = total number of mega-modules tested) were considered significant.

Functional enrichment analysis

To determine if mega-modules with significant overrepresentation of ALSPAC-nominal genes represented an aggregation of functionally related genes, ToppGene (https://toppgene.cchmc.org/) was used to analyze functional enrichment. Categories of biological function, molecular function, cellular component, mouse phenotype, human phenotype, pathways, and drug interaction were tested for over-representation. All genes in the human genome were included in the reference gene set. This set was not limited to the ethanol-responsive genes included in this analysis, in order to preclude functional bias. Significant over-representation results were defined as $p<0.01$ (uncorrected), $n\geq3$ genes overlap and $n\leq1000$ genes per functional group. Given the number of categories and gene sets tested, our discussion below was narrowed to the most relevant categories, defined as Bonferroni-corrected $p<0.1$.

Results

Of the initial 45,037 probesets for the mouse gene expression arrays, 16,131 were associated with human-mouse homologues and had $q_{FDR}<0.1$ for ethanol responsiveness (S-score) in at least one of the three brain regions (Fig 1a). These probesets corresponded to a total of 7,730 genes and were trimmed to a single probeset per gene by filtering for the most abundant probeset as described in Methods. After removing genes that were absent from either the PPI network or the IASPSAD dataset, the final background PPI network for EW-

dmGWAS analysis contained 6,050 genes (nodes) and 30,497 interactions (edges). The nodes contained 25 of the 78 IASPSAD-nominal genes and 24 of the 100 ALSPAC-nominal genes. There was no overlap between the IASPSAD and ALSPAC nominal gene sets.

Prefrontal Cortex

For analysis using PFC expression data for edge-weights, results revealed 3,545 significant modules (q_{FDR}<0.05) containing a total of 4,300 genes, with 14 ALSPAC-nominal genes and 18 IASPSAD-nominal genes. These modules were merged to form 314 mega-modules, all with significant MM-S_n. Twelve mega-modules contained at least one ALSPAC-nominal gene, and 160 contained at least one IASPSAD-nominal gene. However, MM-S_n did not significantly predict mean ALSPAC GWAS gene-wise p-value (β=-0.003, p=0.327, Fig 2).

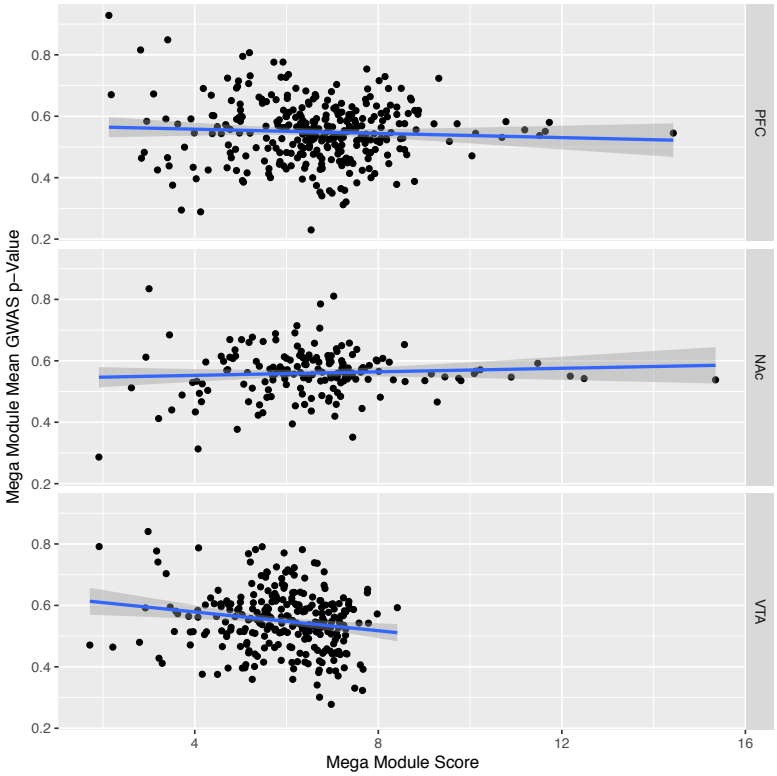


Figure 2: Correlation between each Mega Module's score and average ALSPAC gene-wise GWAS p-value, for the Prefrontal Cortex (PFC) (β=-0.003, p=0.327), Nucleus Accumbens (NAc) (β=0.003, p=0.390), and Ventral Tegmental Area (VTA) (β=-0.02, p=0.003). Blue lines represent the line of best fit, estimated by linear regression, surrounded by their 95% confidence intervals (shaded gray).

Two mega-modules, Aliceblue and Cadetblue, contained multiple ALSPAC-nominal genes (Table 1). Because overrepresentation was tested for 2 mega-modules, $p<0.025$ ($\alpha=0.05/2$) was considered significant. Cadetblue, was significantly overrepresented with ALSPAC-nominal genes (Table 1). Each of Cadetblue's ALSPAC- and IASPSAD-nominal genes was connected to one of its most highly connected hub genes, *ESR1* (estrogen receptor 1; connectivity (k)=31, Eigen-centrality (EC)=1) and *ARRB2* (beta-arrestin-2; k=13, EC=0.25) (Fig 3). Although the ALSPAC-nominal gene overrepresentation was not

significant for Aliceblue, it approached significance (Table 1). Further, Aliceblue had the second-highest MM-S_n in the PFC and contained 3 ALSPAC-nominal genes and 3 IASPSAD-nominal genes (Table 1). For these reasons, Aliceblue was carried through to functional enrichment analysis. Aliceblue's two hub genes were *ELAVL1* ((embryonic lethal, abnormal vision)-like 1; k=165, EC=1) and *CUL3* (cullin 3; k=75, EC=0.21), which were connected to two of the three ALSPAC-nominal genes. Of these, *CPM*'s (carboxypeptidase M's) only edge was with *ELAVL1*, and *EIF5A2*'s (eukaryotic translation initiation factor 5A2's) only edge was with *CUL3* (Fig 3).

Both Cadetblue and Aliceblue showed significant enrichment in several functional categories (S3 Table). In sum, top functional enrichment categories for Aliceblue were related to actin-based movement, cardiac muscle signaling and action, increased triglyceride levels in mice, cell-cell and cell-extracellular matrix adhesion, and syndecan-2-mediated signaling. In contrast, Cadetblue's top enrichment categories involved transcription-regulatory processes, specifically: RNA splicing, chromatin remodeling, protein alkylation and methylation, DNA replication regulation, several immune-related pathways, *NF-κβ* and Wnt signaling pathways, and reductase activity (Tables 2 and 3; S3 Table).

Table 1. ALSPAC Nominal Gene Overrepresentation.

Brain Region	Mega-modules	k_g	MM-S_n	MM-S_n q_{FDR}	Overrep. *p*	Gene	IASPSAD GWAS *p*	ALSPAC GWAS *p*
PFC	aliceblue	392	11.19	<1E-16*	0.063	CPM	0.493	6.48E-05*
						CACNB2	0.978	4.97E-04*
						EIF5A2	0.163	8.06E-04*
						RSL1D1	3.48E-04*	0.217
						SMARCA2	4.91E-04*	0.877
						KIAA1217	8.84E-04*	0.904
	cadetblue	125	6.30	1.08E-06*	0.013*	BCAS2	0.029	4.65E-04*
						PIK3C2A	0.432	9.52E-04*
						RSL1D1	3.48E-04*	0.217
						AKT2	3.90E-05*	0.980
NAc	cadetblue2	195	8.04	8.06E-16*	0.042	CPM	0.493	6.48E-05*
						MGST3	0.358	4.62E-04*
	gray26	12	6.39	9.95E-11*	<0.001*	PCDH7	0.007	2.10E-04*
						BCAS2	0.029	4.65E-04*
VTA	coral	399	4.78	1.00E-06*	0.068	CPM	0.493	6.48E-05*
						DENND2C	0.018	4.33E-04*
						BIRC7	0.930	4.37E-04*
						MGST3	0.358	4.62E-04*
						PIK3CA	7.06E-05*	0.007
						TNN	3.00E-04*	0.018
						ANO6	6.32E-04*	0.780
						SMARCA2	4.91E-04*	0.877
						SIMC1	2.04E-04*	0.977
	limegreen	220	5.22	1.19E-07*	0.054	DENND2C	0.018	4.33E-04*
						EIF5A2	0.163	8.06E-04*
						RSL1D1	3.48E-04*	0.217
						CCND2	1.94E-04*	0.603
						AKT2	3.90E-05*	0.980
	bisque	89	6.22	7.57E-10*	0.006*	ACLY	0.701	2.21E-04*
						PRKG1	0.647	8.26E-04*
						AKT2	3.90E-05*	0.980

The following characteristics are displayed for each mega-module that contained >1 ALSPAC-nominal gene: affiliated brain region; total number of constituent genes (k_g); constituent ALSPAC- and IASPSAD-nominal genes; empirical p-values for ALSPAC-nominal overrepresentation (Overrep. p); MM-S_n, and the associated False Discovery Rate (MM-S_n qFDR). * Significant *p*-values: $p<0.05$ for MM S_n; $p<0.05/n$ for ALSPAC overrepresentation, where n=number of tests per brain region; $p<0.001$ for IASPSAD and ALPSAC GWAS.

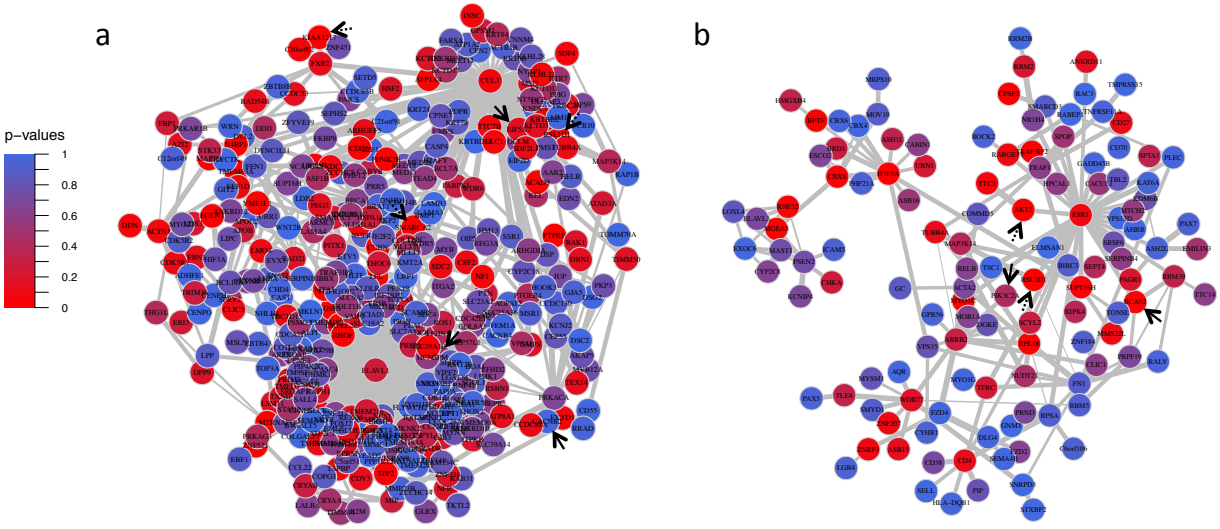


Figure 3: Prefrontal Cortex Mega Modules Cadetblue (a) and Aliceblue (b). Solid black arrows point to ALSPAC GWAS nominal genes, and dotted black arrows represent IASPSAD nominal genes. Edge-width is proportional to the difference in correlation strength between treatment and control mice, and node color represents IASPSAD GWAS p-values.

Table 2. Top Gene Ontology Enrichment Results for PFC Mega Module Cadetblue.

Category	Name	p-value	q-value Bonferroni	Hit Count in Query List	Hit Count in Genome	Hit in Query List
GO: Biological Process	chromatin organization	1.50E-09	4.12E-06	23	776	SMYD1, ESR1, KAT6A, ASH1L, PAGR1, CBX4, KDM6B, ASH2L, MYSM1, PHF21A, BPTF, UBN1, CBX6, SUPT16H, SMARCD3, H3F3B, PAX5, PAX7, BRD1, CABIN1, MGEA5, NR1H4, CBX8
	histone modification	1.97E-06	5.40E-03	14	453	SMYD1, KAT6A, ASH1L, PAGR1, KDM6B, ASH2L, MYSM1, PHF21A, PAX5, PAX7, BRD1, MGEA5, NR1H4, CBX8
	covalent chromatin modification	2.87E-06	7.89E-03	14	468	SMYD1, KAT6A, ASH1L, PAGR1, KDM6B, ASH2L, MYSM1, PHF21A, PAX5, PAX7, BRD1, MGEA5, NR1H4, CBX8
	chromatin remodeling	1.47E-05	4.04E-02	8	165	SMYD1, ESR1, ASH2L, MYSM1, BPTF, SMARCD3, H3F3B, PAX7
	RNA splicing	1.60E-05	4.40E-02	12	403	SRSF6, NUDT21, BCAS2, RBM39, RALY, RBM5, PRPF19, AKT2, CPSF2, SNRPD3, WDR77, AQR
	protein alkylation	2.44E-05	6.71E-02	8	177	SMYD1, ASH1L, ASH2L, PAX5, PAX7, SNRPD3, WDR77, NR1H4
	protein methylation	2.44E-05	6.71E-02	8	177	SMYD1, ASH1L, ASH2L, PAX5, PAX7, SNRPD3, WDR77, NR1H4
GO: Cellular Component	nucleoplasm part	2.23E-05	7.49E-03	16	738	MMS22L, SRSF6, NUDT21, KAT6A, PAGR1, CBX4, ELMSAN1, ASH2L, RBM39, PHF21A, UBN1, TONSL, PRPF19, SPOP, CPSF2, BRD1
	chromosome	1.21E-04	4.07E-02	17	943	MMS22L, PSEN2, BCAS2, ESR1, KAT6A, ASH1L, ZNF207, ASH2L, ESCO2, CBX6, TONSL, SUPT16H, PRPF19, SMARCD3, H3F3B, NR1H4, CBX8
	ribonucleoside-diphosphate reductase complex	1.24E-04	4.17E-02	2	3	RRM2B, RRM2
	DNA replication factor A complex	1.39E-04	4.67E-02	3	16	BCAS2, TONSL, PRPF19
	nuclear replication fork	1.40E-04	4.71E-02	4	41	MMS22L, BCAS2, TONSL, PRPF19
	catalytic step 2 spliceosome	2.96E-04	9.94E-02	5	90	BCAS2, RALY, PRPF19, SNRPD3, AQR
GO: Molecular Function	oxidoreductase activity, acting on CH or CH2 groups	3.32E-05	1.62E-02	3	10	CYP2C8, RRM2B, RRM2
	oxidoreductase activity, acting on CH or CH2 groups, disulfide as acceptor	1.31E-04	6.38E-02	2	3	RRM2B, RRM2
	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	1.31E-04	6.38E-02	2	3	RRM2B, RRM2
	ribonucleoside-diphosphate reductase activity	1.31E-04	6.38E-02	2	3	RRM2B, RRM2
	chromatin binding	1.69E-04	8.24E-02	12	516	ESR1, KAT6A, ASH1L, RELB, CBX4, KDM6B, ASH2L, PHF21A, TLE4, SMARCD3, H3F3B, CABIN1
Mouse Phenotype	increased immunoglobulin level	1.16E-06	2.92E-03	14	307	TRAF3IP2, GADD45B, SEMA4B, PSEN2, ESR1, SPTA1, ASH1L, BIRC3, RELB, MYSM1, CD4, PIK3C2A, RABGEF1, CABIN1
	abnormal humoral immune response	5.52E-06	1.39E-02	18	566	TRAF3IP2, GADD45B, SEMA4B, PSEN2, ESR1, SPTA1, MAP3K14, ASH1L, BIRC3, RELB, TNFRSF11A, MYSM1, CD4, PIK3C2A, CD38, RABGEF1, PAX5, CABIN1
	abnormal immunoglobulin level	7.68E-06	1.93E-02	17	522	TRAF3IP2, GADD45B, SEMA4B, PSEN2, ESR1, SPTA1, MAP3K14, ASH1L, BIRC3, RELB, TNFRSF11A, MYSM1, CD4, PIK3C2A, RABGEF1, PAX5, CABIN1
	increased IgG level	9.35E-06	2.35E-02	11	225	TRAF3IP2, GADD45B, SEMA4B, ESR1, SPTA1, ASH1L, BIRC3, MYSM1, CD4, PIK3C2A, CABIN1
	cortical renal glomerulopathies	1.18E-05	2.96E-02	10	188	TRAF3IP2, GADD45B, PSEN2, MYO1E, ESR1 , SPTA1, RRM2B, ASH1L, RELB, PIK3C2A
	abnormal lymph node morphology	1.85E-05	4.66E-02	14	390	SELL, TRAF3IP2, TRAF1, PSEN2, ESR1, SPTA1, RRM2B, MAP3K14, BIRC3, RELB, TNFRSF11A, CD4, PIK3C2A, PIP
	glomerulonephritis	1.95E-05	4.91E-02	8	121	TRAF3IP2, GADD45B, PSEN2, ESR1, SPTA1, ASH1L, RELB, PIK3C2A
	abnormal B cell physiology	3.21E-05	8.07E-02	18	644	MYO1G, TRAF3IP2, GADD45B, SEMA4B, PSEN2, ESR1, SPTA1, MAP3K14, ASH1L, BIRC3, RELB, TNFRSF11A, MYSM1, CD4, PIK3C2A, RABGEF1, PAX5, CABIN1
Pathway	Signaling by Wnt	2.78E-06	2.47E-03	13	340	LGR4, ASH2L, FZD4, ARRB2, ZNRF3, TLE4, VPS35, H3F3B, AKT2, GNAO1, FZD2, MOV10, RAC3
	NF-kappa B signaling pathway	1.07E-04	9.44E-02	6	95	GADD45B, TRAF1, MAP3K14, BIRC3, RELB, TNFRSF11A
	Apoptosis	1.13E-04	9.97E-02	7	138	GADD45B, TRAF1, SEPT4, SPTA1, MAP3K14, BIRC3, AKT2

Functional enrichment results from ToppFun for Prefrontal Cortex Mega Module Cadetblue, where Bonferroni-corrected $p<0.1$.

Table 3. Top Gene Ontology Enrichment Results for PFC Mega Module Aliceblue.

Category	Name	p-value	q-value Bonferroni	Hit Count in Query List	Hit Count in Genome	Hit in Query List
GO: Biological Process	regulation of actin filament-based movement	4.76E-08	2.07E-04	9	37	FXYD1, ATP1A2, DBN1, GJA5, JUP, KCNJ2, DSC2, DSG2, DSP
	cardiac muscle cell-cardiac muscle cell adhesion	7.53E-08	3.27E-04	5	7	CXADR, JUP, DSC2, DSG2, DSP
	regulation of cardiac muscle cell contraction	1.64E-07	7.11E-04	8	31	FXYD1, ATP1A2, GJA5, JUP, KCNJ2, DSC2, DSG2, DSP
	actin filament-based process	3.57E-07	1.55E-03	36	688	CDC42EP4, ACTN1, MYOZ1, MKLN1, FXYD1, RHOF, SDC4, CUL3, PRR5, CRYAA, ARHGDIA, ATP2C1, CCDC88A, STAU2, DYNLL1, DIXDC1, ATP1A2, CXADR, DBN1, PTGER4, GJA5, JUP, CDK5R1, NF1, KCNJ2, CACNB2, DSC2, DSG2, DSP, ARHGEF5, CASP4, LCP1, CSRP3, LIMK1, LDB3, LRP1
	cell communication involved in cardiac conduction	4.34E-07	1.89E-03	9	47	PRKACA, ATP1A2, CXADR, GJA5, JUP, CACNB2, DSC2, DSG2, DSP
	desmosome organization	8.59E-07	3.73E-03	5	10	SNAI2, JUP, DSG2, DSP, PKP3
	cardiac muscle cell action potential	1.07E-06	4.65E-03	9	52	ATP1A2, CXADR, GJA5, JUP, KCNJ2, CACNB2, DSC2, DSG2, DSP
	cardiac muscle cell contraction	1.07E-06	4.65E-03	9	52	FXYD1, ATP1A2, GJA5, JUP, KCNJ2, CACNB2, DSC2, DSG2, DSP
	bundle of His cell to Purkinje myocyte communication	1.55E-06	6.72E-03	5	11	GJA5, JUP, DSC2, DSG2, DSP
	regulation of cardiac muscle cell action potential	2.30E-06	9.99E-03	6	20	CXADR, GJA5, JUP, DSC2, DSG2, DSP
	bundle of His cell-Purkinje myocyte adhesion involved in cell communication	2.63E-06	1.14E-02	4	6	JUP, DSC2, DSG2, DSP
	regulation of heart rate by cardiac conduction	2.65E-06	1.15E-02	7	31	GJA5, JUP, KCNJ2, CACNB2, DSC2, DSG2, DSP
	cardiac conduction	3.37E-06	1.46E-02	13	131	FXYD1, PRKACA, ATP1A2, ATP1A4, CXADR, GJA5, JUP, KCNJ2, CACNB2, CACNB4, DSC2, DSG2, DSP
	cardiac muscle cell action potential involved in contraction	7.69E-06	3.34E-02	7	36	GJA5, JUP, KCNJ2, CACNB2, DSC2, DSG2, DSP
	regulation of actin filament-based process	1.05E-05	4.58E-02	21	343	CDC42EP4, FXYD1, SDC4, ARHGDIA, CCDC88A, STAU2, DIXDC1, ATP1A2, DBN1, PTGER4, GJA5, JUP, CDK5R1, KCNJ2, DSC2, DSG2, DSP, ARHGEF5, CSRP3, LIMK1, LRP1
	lipoprotein localization	1.34E-05	5.83E-02	5	16	APOB, APOC2, MSR1, CUBN, LRP1
	lipoprotein transport	1.34E-05	5.83E-02	5	16	APOB, APOC2, MSR1, CUBN, LRP1
	regulation of cardiac muscle contraction	1.36E-05	5.91E-02	9	70	FXYD1, PRKACA, ATP1A2, GJA5, JUP, KCNJ2, DSC2, DSG2, DSP
GO: Cellular Component	intercalated disc	2.90E-06	1.53E-03	9	59	ACTN1, ATP1A2, CXADR, GJA5, JUP, KCNJ2, DSC2, DSG2, DSP
	cell-cell contact zone	1.56E-05	8.21E-03	9	72	ACTN1, ATP1A2, CXADR, GJA5, JUP, KCNJ2, DSC2, DSG2, DSP
	desmosome	1.61E-04	8.49E-02	5	26	JUP, DSC2, DSG2, DSP, PKP3
GO: Molecular Function	protein binding involved in heterotypic cell-cell adhesion	8.62E-07	7.88E-04	5	10	CXADR, JUP, DSC2, DSG2, DSP
	protein binding involved in cell adhesion	1.15E-06	1.05E-03	6	18	CXADR, ITGA2, JUP, DSC2, DSG2, DSP
	protein binding involved in cell-cell adhesion	2.62E-06	2.39E-03	5	12	CXADR, JUP, DSC2, DSG2, DSP
	cell adhesive protein binding involved in bundle of His cell-Purkinje myocyte communication	2.64E-06	2.41E-03	4	6	JUP, DSC2, DSG2, DSP
Human Phenotype	Dilated cardiomyopathy	4.35E-05	3.89E-02	9	87	ACAD9, CRYAB, UBR1, JUP, DSG2, DSP, LAMA4, CSRP3, LDB3
	Right ventricular cardiomyopathy	8.82E-05	7.90E-02	4	13	JUP, DSC2, DSG2, DSP
Mouse Phenotype	increased circulating triglyceride level	1.27E-05	4.77E-02	16	179	ALPI, COL1A1, VLDLR, AGPAT2, WRN, APOB, APOC2, TXNIP, RSBN1, CSF2, PRKACA, BGLAP, MED13, LEPR, LIPC, LRP1
Pathway	Non-integrin membrane-ECM interactions	3.41E-05	4.72E-02	7	46	ACTN1, SDC2, SDC4, ITGA2, LAMA3, LAMA4, LAMB3
	Syndecan-2-mediated signaling events	4.44E-05	6.14E-02	6	33	SDC2, CSF2, PRKACA, ITGA2, NF1, LAMA3

Functional enrichment results from ToppFun for Prefrontal Cortex Mega Module Aliceblue, where Bonferroni-corrected p<0.1.

Nucleus Accumbens

Using NAc acute ethanol expression data for edge-weights yielded 3,460 significant modules containing a total of 4,213 genes, 15 of which were ALSPAC-nominal and 16 of which were IASPSAD-nominal. After merging by content similarity, there were 171 significant mega-modules. Nineteen MM contained at least one ALSPAC-nominal gene, and 73 MM contained at least one IASPSAD-nominal gene. However, MM S_n did not significantly predict MM mean ALSPAC GWAS gene-wise p-value (β=0.003, p=0.390). Two MMs, Cadetblue2 and Gray26, each contained two ALSPAC-nominal genes (Table 1). Because there were 2 tests for overrepresentation, p<0.025 (α=0.05/2) was considered significant. Gray26, was significantly overrepresented with ALSPAC-nominal genes, and Cadetblue2 showed a trend towards overrepresentation with significance before correcting for multiple testing (Table 1).

Gray26's most central hub gene was *HNRNPU* (heterogeneous nuclear ribonucleoprotein U; connectivity=6, Eigen-centrality=1), followed by *RBM39* (RNA binding motif protein 39; k=3, EC=0.46) and *CSNK1A1* (k=3, EC=0.37). The two ALSPAC-nominal genes *BCAS2* (breast carcinoma amplified sequence 2) and *PCDH7* (protocadherin 7), shared their only edges with *RBM39* and *HNRPNPU*, respectively (Fig 4a). As seen in the PFC's Aliceblue, *ELAVL1* was a hub gene of Cadetblue2. *ELAVL1* (k=136, EC=1) was connected to both of the ALSPAC-nominal genes, and served as the only connection for *CPM* and one of two connections for *MGST3* (microsomal glutathione S-transferase 3) (Fig 4b). Strikingly, PFC Aliceblue and NAc Cadetblue 2 showed a highly significant overlap in their gene content, with 72 overlapping genes (S2 Table; $p=2.2 \times 10^{-16}$).

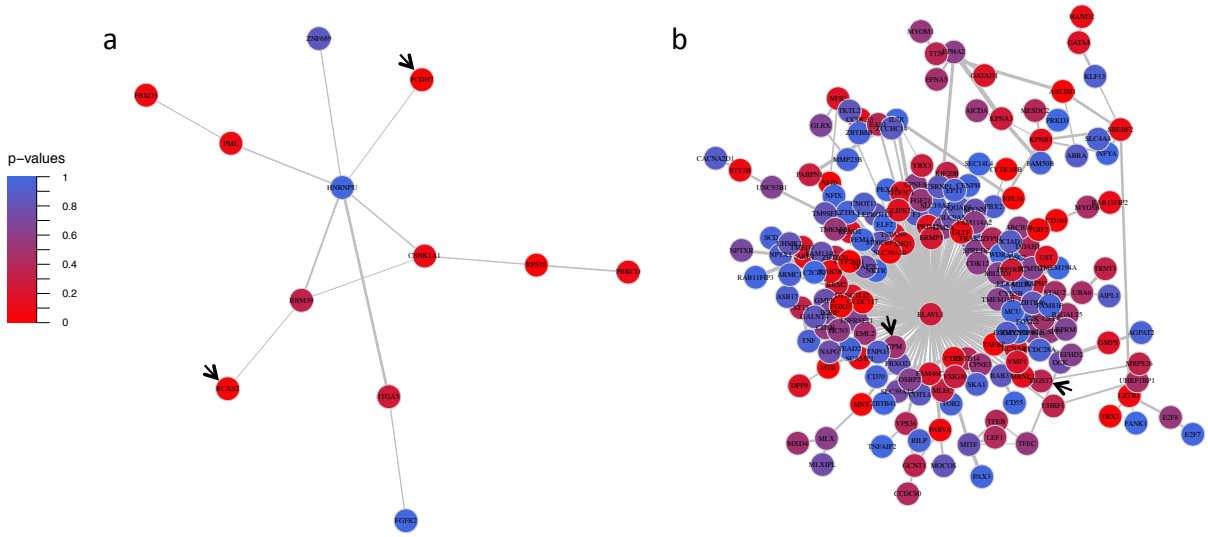


Figure 4: Nucleus Accumbens Mega Modules Gray26 (a) and Cadetblue2 (b). Solid black arrows point to ALSPAC GWAS nominal genes. These modules did not contain IASPSAD nominal genes. Edge-width is proportional to the difference in correlation strength treatment and control mice, and node color represents IASPSAD GWAS p-values.

Both Cadetblue2 and Gray26 were significantly enriched with several functional categories (S3 Table). Like PFC Cadetblue, NAc Cadetblue2 was functionally enriched for gene groups related to nuclear function with transcription regulation pathways, particularly those involving RNA polymerase activity. Gray26 was most significantly enriched with genes related to functions involving: telomere maintenance, organelle organization, ribonucleoprotein complexes, and syndecan-mediated signaling (Tables 4 and 5; S3 Table).

Table 4. Top Gene Ontology Enrichment Results for Nucleus Accumbens Mega Module Cadetblue2.

Category	Name	p-value	q-value Bonferroni	Hit Count in Query List	Hit Count in Genome	Hit in Query List
GO: Biological Process	negative regulation of transcription from RNA polymerase II promoter	9.38E-06	2.93E-02	23	810	TGIF2, ZBTB20, SREBF2, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, MNT, TBX2, MLX, YBX3, TFAP2C, MXD4, E2F8, ZBTB14, MLXIPL, UHRF1, TNF, ELK4, PAX3, LEF1
GO: Molecular Function	RNA polymerase II transcription factor activity, sequence-specific DNA binding	1.80E-09	1.20E-06	27	678	ZBTB20, SREBF2, GATA4, E2F7, CSRNP1, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, TEAD2, MLX, YBX3, FOXJ3, TFAP2C, E2F8, MLXIPL, KLF13, ELF2, ELK4, PAX3, LEF1
	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific binding	3.04E-06	2.03E-03	11	182	ZBTB20, SREBF2, E2F7, MITF, MNT, TBX2, MLX, YBX3, TFAP2C, E2F8, MLXIPL
	transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding	6.11E-06	4.08E-03	15	365	ZBTB20, SREBF2, FOXL2, NFIB, NFIC, MITF, NFYA, HAND2, TBX2, TFEB, TFAP2C, E2F8, MLXIPL, KLF13, LEF1
	RNA polymerase II regulatory region sequence-specific DNA binding	8.95E-06	5.98E-03	20	632	SREBF2, GATA4, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, MLX, YBX3, TFAP2C, E2F8, MLXIPL, KLF13, LEF1
	transcription regulatory region DNA binding	9.52E-06	6.36E-03	24	862	SREBF2, GATA4, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, MLX, YBX3, TFAP2C, E2F8, ZBTB14, MLXIPL, KLF13, UHRF1, TNF, ELK4, LEF1
	regulatory region DNA binding	1.01E-05	6.74E-03	24	865	SREBF2, GATA4, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, MLX, YBX3, TFAP2C, E2F8, ZBTB14, MLXIPL, KLF13, UHRF1, TNF, ELK4, LEF1
	RNA polymerase II regulatory region DNA binding	1.03E-05	6.87E-03	20	638	SREBF2, GATA4, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, MLX, YBX3, TFAP2C, E2F8, MLXIPL, KLF13, LEF1
	regulatory region nucleic acid binding	1.07E-05	7.14E-03	24	868	SREBF2, GATA4, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, MLX, YBX3, TFAP2C, E2F8, ZBTB14, MLXIPL, KLF13, UHRF1, TNF, ELK4, LEF1
	transcription regulatory region sequence-specific DNA binding	1.32E-05	8.82E-03	21	705	SREBF2, GATA4, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, MLX, YBX3, TFAP2C, E2F8, MLXIPL, KLF13, UHRF1, LEF1
	sequence-specific double-stranded DNA binding	2.50E-05	1.67E-02	21	736	SREBF2, GATA4, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, MLX, YBX3, TFAP2C, E2F8, MLXIPL, KLF13, UHRF1, LEF1
	core promoter proximal region sequence-specific DNA binding	7.08E-05	4.73E-02	14	399	SREBF2, GATA4, FOXL2, NFIB, NFIC, MITF, NFYA, TBX2, TFEB, E2F8, MLXIPL, KLF13, UHRF1, LEF1
	core promoter proximal region DNA binding	7.47E-05	4.99E-02	14	401	SREBF2, GATA4, FOXL2, NFIB, NFIC, MITF, NFYA, TBX2, TFEB, E2F8, MLXIPL, KLF13, UHRF1, LEF1
	transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	9.15E-05	6.11E-02	13	358	GATA4, CSRNP1, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, HAND2, TFEB, TFAP2C, KLF13, LEF1
	double-stranded DNA binding	1.25E-04	8.37E-02	21	824	SREBF2, GATA4, E2F7, FOXL2, NFIB, NFIC, NFIX, MITF, NFYA, MNT, HAND2, TBX2, TFEB, MLX, YBX3, TFAP2C, E2F8, MLXIPL, KLF13, UHRF1, LEF1
Human Phenotype	Synophrys	3.61E-05	2.06E-02	5	48	ZBTB20, NFIX, MITF, KLF13, PAX3
Mouse Phenotype	absent coat pigmentation	2.38E-05	6.28E-02	4	15	MITF, TFEB, TFEC, PAX3

Functional enrichment results from ToppFun for Nucleus Accumbens Mega Module Cadetblue2, where Bonferroni-corrected $p<0.1$.

Table 5. Top Gene Ontology Enrichment Results for Nucleus Accumbens Mega Module Gray26.

Category	Name	p-value	q-value Bonferroni	Hit Count in Query List	Hit Count in Genome	Hit in Query List
GO: Biological Process	negative regulation of telomere maintenance via telomerase	2.46E-05	2.92E-02	2	12	HNRNPU, PML
	negative regulation of organelle organization	4.65E-05	5.52E-02	4	340	PRKCD, FGFR2, HNRNPU, PML
	negative regulation of telomere maintenance via telomere lengthening	5.06E-05	6.00E-02	2	17	HNRNPU, PML
GO: Cellular Component	ribonucleoprotein complex	8.99E-04	8.99E-02	4	751	CSNK1A1, RPS18, BCAS2, HNRNPU
	intracellular ribonucleoprotein complex	8.99E-04	8.99E-02	4	751	CSNK1A1, RPS18, BCAS2, HNRNPU
Pathway	Syndecan-4-mediated signaling events	2.67E-04	7.44E-02	2	31	PRKCD, ITGA5
	Syndecan-2-mediated signaling events	3.03E-04	8.44E-02	2	33	PRKCD, ITGA5

Functional enrichment results from ToppFun for Nucleus Accumbens Mega Module Gray26, where Bonferroni-corrected p<0.1.

Ventral Tegmental Area

Use of VTA control/ethanol gene expression responses for edge weighting initially resulted in 3,519 significant modules containing a total of 4,188 genes in EW-dmGWAS analysis. Merging by content similarity, resulted in 276 MMs, each with a significant MM S_n. Seventeen ALSPAC-nominal genes and 19 IASPSAD-nominal genes were spread across 25 and 156 mega-modules, respectively. Furthermore, MM-S_n significantly predicted mean ALSPAC GWAS gene-wise *p*-value (β=-0.02, p=0.003).

Mega-modules with the highest representation of ALSPAC-nominal genes included Coral, Limegreen, and Bisque (Table 1). Because there were 3 tests for overrepresentation, p<0.017 (α=0.05/3) was considered significant. Although overrepresentation of ALSPAC-nominal genes was not significant in Coral and Limegreen, it was significant in Bisque, which has the highest MM-S_n of the three (Table 1; Fig 5). Bisque contained four highly interconnected genes: *USP21* (ubiquitin specific peptidase 21; k=10, EC=1), *USP15* (ubiquitin specific peptidase 15; k=10, EC=0.65), *TRIM25* (tripartite motif-containing 25;

k=10, EC=0.49), and *HECW2* (HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2; k=12, EC=0.48). *HECW2* and *TRIM25* shared edges with this MM's IASPSAD-nominal genes *PRKG1* (protein kinase, cGMP-dependent, type I) and *ACLY* (ATP citrate lyase), respectively. However, none of the hub genes shared an edge with Bisque's ALSPAC nominal gene, *AKT2* (AKT serine/threonine kinase 2). Finally, Bisque had significant enrichment in several functional categories (S3 Table). It was most significantly enriched with genes associated with ubiquitination, ligase and helicase activity, and eukaryotic translation elongation (Table 6; S3 Table).

Table 6. Top Gene Ontology Enrichment Results for Ventral Tegmental Area Mega Module Bisque.

Category	Name	p-value	q-value Bonferroni	Hit Count in Query List	Hit Count in Genome	Hit in Query List
GO: Cellular Component	nucleolus	6.41E-07	1.24E-04	17	894	ZNF106, NEK2, EEF1D, RPL36, PNKP, SELENBP1, ZNF655, RPS9, WRN, GATA3, ZFHX3, RORC, DGCR8, TTC3, ARNTL2, NEK11, RPL18
	eukaryotic translation elongation factor 1 complex	1.27E-04	2.47E-02	2	4	EEF1D, EEF1A2
GO: Molecular Function	ubiquitin-protein transferase activity	4.98E-07	1.33E-04	12	414	RC3H2, TRAF4, UBE2K, TRIM2, TRIM25, TRIM9, HECW2, TRIM8, UBE2S, RNF114, TTC3, TRIM37
	ubiquitin-like protein transferase activity	9.70E-07	2.59E-04	12	441	RC3H2, TRAF4, UBE2K, TRIM2, TRIM25, TRIM9, HECW2, TRIM8, UBE2S, RNF114, TTC3, TRIM37
	acid-amino acid ligase activity	3.42E-06	9.12E-04	9	259	RC3H2, TRIM2, TRIM25, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	ligase activity, forming carbon-nitrogen bonds	9.78E-06	2.61E-03	9	295	RC3H2, TRIM2, TRIM25, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	tubulin-glycine ligase activity	1.87E-05	5.00E-03	8	244	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	protein-glycine ligase activity	1.87E-05	5.00E-03	8	244	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	protein-glycine ligase activity, initiating	1.87E-05	5.00E-03	8	244	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	coenzyme F420-0 gamma-glutamyl ligase activity	1.87E-05	5.00E-03	8	244	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	ribosomal S6-glutamic acid ligase activity	1.87E-05	5.00E-03	8	244	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	coenzyme F420-2 alpha-glutamyl ligase activity	1.87E-05	5.00E-03	8	244	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-alanine ligase activity	1.87E-05	5.00E-03	8	244	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	protein-glycine ligase activity, elongating	1.87E-05	5.00E-03	8	244	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	tubulin-glutamic acid ligase activity	2.05E-05	5.46E-03	8	247	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	protein-glutamic acid ligase activity	2.17E-05	5.79E-03	8	249	RC3H2, TRIM2, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	ligase activity	2.38E-05	6.35E-03	10	415	LIG3, RC3H2, TRIM2, TRIM25, TRIM9, HECW2, TRIM8, RNF114, TTC3, TRIM37
	DNA helicase activity	2.43E-04	6.49E-02	4	65	ERCC2, GTF2H4, RAD54B, WRN
Pathway	Eukaryotic Translation Elongation	1.67E-04	8.37E-02	5	98	EEF1D, RPL36, RPS9, EEF1A2, RPL18

Functional enrichment results from ToppFun for Ventral Tegmental Area Mega Module Bisque, where Bonferroni-corrected $p<0.1$.

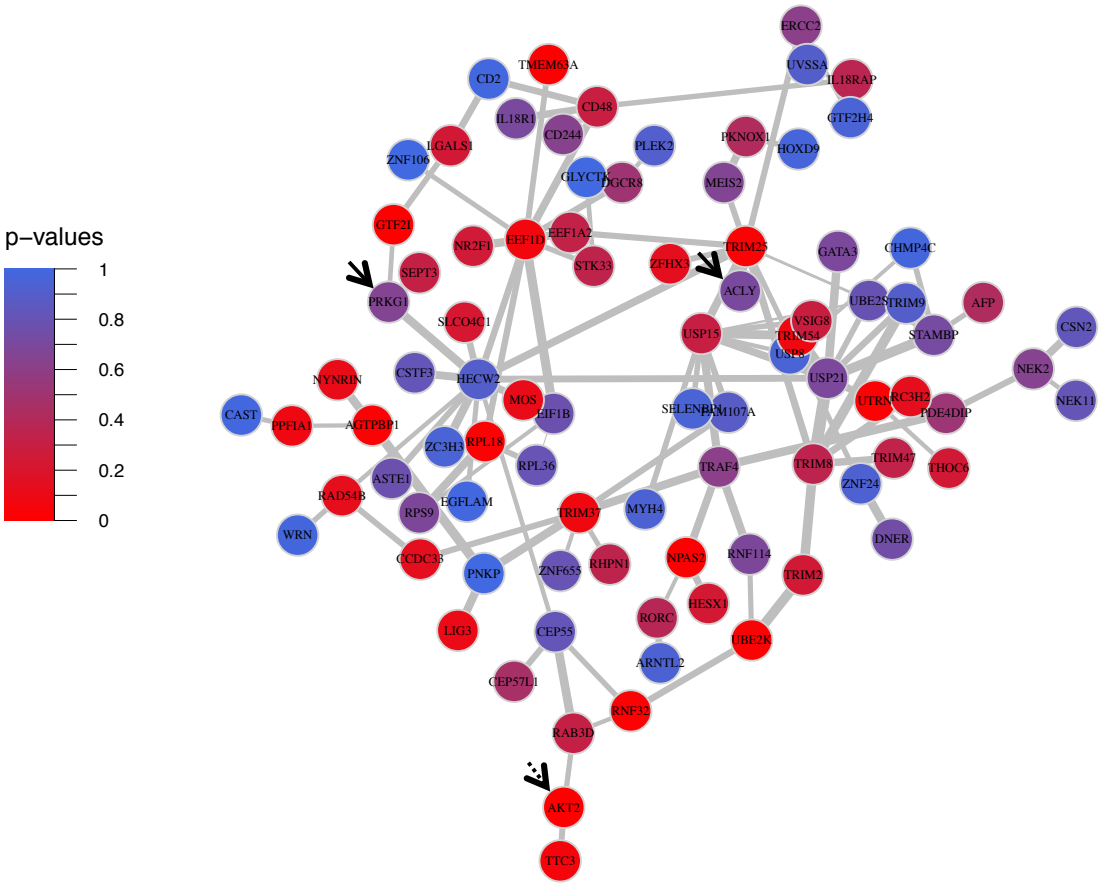


Figure 5: Ventral Tegmental Area Mega Modules Bisque. Solid black arrows point to ALSPAC GWAS nominal genes, and dotted black arrows represent IASPSAD nominal genes. Edge-width is proportional to the difference in correlation strength between treatment and control mice, and node color represents IASPSAD GWAS p-values.

Discussion

To our knowledge, this is the first study to directly co-analyze human GWAS with mouse brain ethanol-responsive gene expression data to identify ethanol-related gene networks relevant to AD. Unlike previous studies that have employed cross-species validation methods for specific genes or gene sets, this study analyzed human and mouse data in tandem to identify gene networks across the entire genome, using the EW-dmGWAS algorithm. This approach successfully identified significantly ethanol-responsive and AD-associated gene networks, or modules. We further improved the existing EW-dmGWAS algorithm by merging highly redundant modules to create more parsimonious mega-modules, thus decreasing complexity without sacrificing significance. Additionally, we validated these results by testing for overrepresentation with, and mega-module score prediction by, signals from an independent GWAS dataset. Overall, our findings suggest that such direct integration of model organism expression data with human protein interaction and GWAS data can productively leverage these data sources. Furthermore, we present initial evidence for novel, cross-validated gene networks warranting further study for mechanisms underlying AUD.

Identification of Network-Level Associations Across GWAS Datasets

One major concern with existing GWASs on AD had been the relative lack of replication across studies. Although some very large GWASs on alcohol consumption have shown replicable results (23, 26, 60), those do not account for all previously identified associations. We reasoned that our integrative gene network-querying approach might identify networks that shared signals from different GWASs on AD, even if the signals were

not from the same genes across GWASs. Concordant with this hypothesis, VTA mega-module scores significantly predicted average gene-wise p-values from an independent GWAS dataset, ALSPAC (Fig 2). This suggests that ethanol-responsive gene expression networks in this brain region may be particularly sensitive to genetic variance and thus are highly relevant to mechanisms contributing to genetic risk for AD. This is possibly attributable to the involvement of VTA dopaminergic reward pathways in the development of AD (15). Further investigation of dopaminergic neuronal response to acute ethanol administration, and the association between this response and proclivity for developing dependence is needed.

Although scores did not prioritize mega-modules with respect to ALSPAC results in PFC and NAc, individual mega-modules were overrepresented with ALSPAC signals (Table 1). The ALSPAC-overrepresented VTA and PFC mega-modules also contained nominally significant genes from the GWAS dataset used for the network analysis, IASPSAD. These results suggest that the integration of acute ethanol-related expression data from mice and human PPI can identify functional networks that associate signals from different GWAS datasets.

Composition and Structure of Mega-Modules

Functional composition of mega-modules varied between brain regions for the most part. For example, although Aliceblue (PFC) and Cadetblue2 (NAc) shared the hub gene *ELAVL1*, ALSPAC-nominal gene *CPM*, and had a significant overlap in their gene content, their functional enrichment results were very different (Tables 3 and 4). These results suggest that brain regional ethanol-responsive gene expression results likely had an important impact

on composition of networks, thus leveraging protein-protein interaction network information and GWAS results.

Despite such differences, the mega-modules presented in Table 1 shared certain structural similarities. Most of the IAPSAD- and ALSPAC-nominal genes in these modules shared edges with hub genes (Fig 3-5). These hub genes included: *CUL3 and ELAVL1* from PFC Aliceblue; *ESR1* from PFC Cadetblue; *ELAVL1*rom NAc Cadetblue2; *TRIM25* and *HECW2* from VTA Bisque. Further, GWAS nominally significant genes (IASPAD or ALSPAC) generally were not hub genes in the derived networks (see Fig 3-5; S2 Table). This may be consistent with the general tenet that genetic variation in complex traits does not produce major alterations in cellular function, but rather modulation of cellular mechanisms for maintaining homeostasis. Hub genes may be more closely functionally related to a given trait, but likely have such widespread influence so as to be evolutionarily resistant to genetic variation in complex traits. This is also consistent with the hypothesis that omnigenic influences are an important feature of complex traits such as AUD (132).

One hub gene was found to influence network structure in both PFC and NAc. *ELAVL1* is a broadly expressed gene that acts as a RNA-binding protein in AU-rich domains, generally localized within 3'-UTRs of mRNA. As such, *ELAVL1* has been shown to alter mRNA stability by altering binding of miRNA or other factors influencing mRNA degradation (133) and has been implicated in activity-dependent regulation of gene expression in the brain with drug abuse (134). The large interaction space for *ELAVL1* in PFC Alice Blue and NAc Cadetblue 2 and the multiple nominal GWAS hits within these genes suggest that *ELAVL1* could have an important modulatory function on the network of genes susceptible to genetic variation in AUD.

Functional Aspects of Mega-Modules

This theory regarding network structure is further supported by our functional enrichment analysis, which revealed several small groups of functionally related genes within each mega-module. All of the mega-modules discussed above (Table 1) contained at least one GWAS-nominal gene in the top enrichment groups, except Cadetblue2, which still had GWAS-nominal genes in its significant enrichment groups (S3 Table).

Another unifying feature across these mega-modules, except Aliceblue, was significant functional enrichment for pathways that regulate gene expression. Specifically, these pathways were related to chromatin organization, RNA splicing, and translation- and transcription-related processes (S3 Table). This is not surprising, as alterations in gene expression have long been proposed as a mechanism underlying long-term neuroplasticity resulting in ethanol-dependent behavioral changes, and eventually dependence (135).

In contrast, the largest functional enrichment groups unique to Aliceblue were related to actin-based filaments and cardiac function (Tables 2 and 3). Actin not only provides cytoskeletal structure to neurons, but also functions in dendritic remodeling in neuronal plasticity, which likely contributes to AD development (136, 137). Aliceblue was also significantly enriched for the syndecan-2 signaling pathway, and contained the *SDC2* gene itself, which functions in dendritic structural changes together with F-actin (138). Additionally, the most significant enrichment group unique to Cadetblue was the Wnt signaling pathway, which also regulates actin function (139, 140). Of note, a prior study has shown that *ARRB2* (a Cadetblue hub gene and member of Wnt signaling pathway) knockout rats display significantly decreased levels of voluntary ethanol consumption and psychomotor stimulation in response to ethanol (141). These findings highlight the potential

importance of postsynaptic actin-related signaling and dendritic plasticity in PFC gene networks responding to acute ethanol and contributing to genetic risk for AD. Future studies may aim to confirm this association by investigating changes in actin and dendritic processes in response to acute ethanol exposure, and whether or not the degree of these changes is associated with development of dependence.

Finally, although the NAc Cadetblue2 mega-module was highly enriched for functions related to transcriptional regulation, it also contained the gene *FGF21* within its interaction space (S2 Table and Fig 4b). FGF21 is a member of the fibroblast growth factor gene family and is a macronutrient responsive gene largely expressed in liver. Importantly FGF21 has been shown to be released from the liver by ethanol consumption and negatively regulates ethanol consumption by interaction with brain FGF-receptor/beta-Klotho complexes. Beta-Klotho, a product of the *KLB* gene, is an obligate partner of the FGF receptor and has recently been shown to have a highly significant association with alcohol consumption in recent very large GWAS studies (26, 60). Although the role of *FGF21* and *KLB* in AD are not currently known, the association of *FGF21* with the Cadetblue2 mega-module, containing nominally associated genes from AD GWAS studies, is a possible additional validation of the utility of our studies integrating protein-protein interaction information (tissue non-specific), AD GWAS (tissue non-specific) and brain ethanol-responsive gene expression.

Potential Weaknesses and Future Studies

The studies presented here provide evidence for the utility of integrating genomic expression data with protein-protein interaction networks and GWAS data in order to gain a

better understanding of the genetic architecture of complex traits, such as AD. Our analysis also generated several testable hypotheses regarding gene networks and signaling mechanisms related to ethanol action and genetic burden for AD. However, these studies utilized acute ethanol-related expression data in attempting to identify mechanisms of AD, a chronic ethanol exposure disease. Use of a chronic exposure model could provide for a more robust integration of the expression data and GWAS signals. However, we feel the current study is valid, since acute responses to ethanol have been repeatedly shown to be a heritable risk factor for AD (142-144). Further, large GWAS studies have recently shown significant genetic correlation and overlapping significant genes between alcohol consumption and alcohol dependence phenotypes (145, 146). We have also recently demonstrated a very high degree of overlap in mouse brain expression changes between acute ethanol exposure and a chronic ethanol vapor exposure model thought to mimic aspects of alcohol dependence (97). Our laboratory has also recently reported that an acute ethanol-responsive gene network from the same microarray data used for studies in this manuscript showed significant association, at a network level, with AD in data from the COGA GWAS analysis of AD (122). Finally, the cross-species analysis of acute ethanol responses and AD allowed us to explore networks involved in specific brain regional initial response to ethanol that are also related to dependence. Therefore, our findings may have implications for mechanistic activations or changes occurring upon initial ethanol exposure, and ultimately contributing to the development of dependence.

A potential shortcoming for this work regards the limited size of the GWAS studies utilized and differences in phenotypic assessment. The IASPSAD study was based on AD diagnosis, whereas ALSPAC was based on a symptom factor score. Had we used larger

GWAS studies based on the same assessment criteria, it is possible that greater overlap of GWAS signals within mega-modules would have been observed. Recent large GWAS studies on ethanol have, to date, generally concerned measures of ethanol consumption, rather than a diagnosis of alcohol dependence per se (26, 60). For this reason, we focused this initial effort on GWAS studies concerned with alcohol dependence. However, using the IASPAD and ALSPAC studies allowed us to identify gene networks that are robust across both the severe end of the phenotypic spectrum (i.e. diagnosable AD), and for symptoms at the sub-diagnostic level.

Finally, although gene expression correlations are relatively largely conserved across mice and humans (105), network structure can vary across time and ethanol exposure (147) (28), as can the associations between genetic variants and gene expression (i.e. eQTL). This study examines networks and eQTL across low and high drinkers at one specific time point (24hrs after last ethanol access period), which could fail to represent networks and contributing loci unique to each drinking group or to other post-exposure time points. Networks and their respective eQTL are being examined within each of these drinking groups, in ongoing analyses in our laboratory. However, power to detect network eQTL will be decreased for eQTL that are common between low and high drinkers, due to the decreased sample size (N=100 per group, as opposed to N=200 total), hence the benefits of performing these analyses on all 200 ethanol-exposed mice. Future studies should examine gene expression networks in the PFC and their eQTL in larger samples of high and low voluntarily drinking mice, and collect brain tissue at various time points after (and possibly during) the last period of ethanol exposure.

Overall, this analysis successfully identified novel ethanol-responsive, AD-associated, functionally enriched gene expression networks in the brain that may play a causal role in the developmental pathway from first ethanol exposure to AD. This is the first analysis to identify such networks by directly co-analyzing brain regional gene expression data, protein-protein interaction data, and GWAS summary statistics. The identified modules provided insight into common pathways between differing signals from independent, largely underpowered, yet deeply phenotyped GWAS datasets. This supports the conjecture that the integration of different GWAS results at a gene network level, rather than simply looking for replication of individual gene signals, could make use of previously underpowered datasets and identify common genetic mechanisms relevant to AD. Future expansion of such approaches to integrate additional model organism chronic ethanol-responsive gene sets with the rapidly evolving GWAS literature on alcohol consumption and dependence, together with validation of key targets by gene targeting in animals models, may provide both novel insight for the neurobiology of AD and the development of improved therapeutic approaches.

Chapter 3

Characterization of Voluntary Ethanol Consumption in 8 Inbred Mouse Strains with Chronic versus Intermittent Ethanol Access

Introduction

Model organism genetic studies, particularly in mice, have thus been used extensively to identify loci/genes associated with behavioral responses to ethanol. Furthermore, as human genome-wide association studies on AUD become larger and produce more significant findings, there will be a need to compare these to model organism findings to validate candidate genes and decipher biological mechanisms. However, mouse forward genetic studies on ethanol behaviors to date have produced few candidate genes validated in humans. This has largely been due to several factors, including: possible inaccurate modeling of human abusive alcohol consumption, limited genetic diversity and limited mapping resolution.

Commonly used inbred mouse lines lack allelic diversity compared to humans. Furthermore, common laboratory mouse strains possess many large DNA blocks of linkage disequilibrium (LD), often showing extensive cross-strain homology, due to common ancestries and lack of sufficient meiotic events to produce recombination. Large LD blocks and allelic homogeneity preclude the identification of novel alcohol behavior-associated variants, and the narrowing of trait-associated regions to identify candidate genes. Recent efforts at increasing allelic diversity and mapping precision in mouse models, however, have used complex breeding schemes of multiple traditional inbred lines and incorporation of wild-derived strains, which harbor considerable genetic diversity (35).

The Diversity Outbred (DO) panel, in particular, was created specifically for genetic analysis of complex behavioral traits. DO mice originated from interbreeding mice from the earlier generations of inbreeding in the creation of the Collaborative Cross lines. These mouse lines have great allelic heterogeneity resulting from 144 different breeding patterns

involving 8 progenitor strains: C57BL/6J, A/J, 129S1/SvlmJ, NOD/ShiLtJ, NZO/HiLtJ, CAST/EiJ, WSB/EiJ, and PWK/PhJ (35). The latter 3 strains are wild-derived lab strains, whose "wild" alleles contribute greatly and uniquely to both genetic and behavioral variation in the DO mice (32, 90). DO mice have already afforded identification of behavioral trait-related loci (35). However, to our knowledge, research on alcohol-related traits in DO mice has not yet been published.

Recent progress has also been made in the development of experimental designs providing possible increased consilience for studying AUD. Intermittent ethanol access (IEA) paradigms, comprised of alternating periods of free-choice ethanol access and forced abstinence, produce higher levels of ethanol consumption than continuous access models and have been shown to mimic aspects of compulsive ethanol consumption seen in humans (36, 37, 77, 79, 148). IEA paradigms model the progressive increase in ethanol drinking seen in humans during the early stages of development of Alcohol Use Disorder (AUD) (77). Prior studies have identified strain-dependent variance in IEA, as seen with ethanol consumption and preference under chronic access paradigms, suggesting genetic contributions to variance in IEA (77-79).

However, to our knowledge, use of DO mice for studies on ethanol-related traits such as IEA consumption has not been reported. Prior to pursuit of such complex analyses in DO mice, we hypothesized that studies on the heritability and diversity of constant vs. IEA consumption in the DO progenitor lines would determine the feasibility and aid design of an extensive genetic study in DO mice. Therefore, here we assess voluntary ethanol consumption and preference in the DO progenitor strains, to contrast both constant access and IEA consumption variation across these strains. To our knowledge, this study is the first

to characterize voluntary ethanol intake and preference in several of these strains (specifically NZO/HiLtJ, CAST/EiJ, WSB/EiJ, and PWK/PhJ). Our results show remarkable diversity in ethanol consumption patterns across the progenitor strains with IEA having strong strain-dependent contrasts with the constant ethanol access paradigm. The large heritability for these traits across the progenitor strains suggests that future large scale genetic screens for ethanol consumption in the DO mice are both feasible and likely to provide novel understanding of molecular factors influencing progression of ethanol consumption as seen in AUD.

Materials and Methods

Ethics Statement

All animal care and euthanasia procedures were performed in accordance with the requirements and recommendations presented by the United States Department of Agriculture Animal Welfare Act and Regulations, Public Health Services Policy on Humane Care and Use of Laboratory Animals, and American Association for Accreditation of Laboratory Animal Care**. Humane endpoints were established by the same standards.

Animals

All mice were obtained from Jackson Laboratories at 8 weeks of age and were initially habituated to the animal vivarium for 1-2 weeks in group housing (n=4) before then being habituated to single housing for 1 week prior to experimentation. Mice were housed with 12-hour light-dark cycles and ad libitum access to water and standard chow (#7912, Harlan Teklad, Madison, WI, United States), in humidity-and temperature-controlled rooms with shaved cedar bedding. Cages were changed and body weights were recorded weekly.

Due to availability of strains from the supplier, two separate experiments were run for complete analysis of all progenitor strains. Experiment 1 contained 10 mice from each of the following strains: 1) C57BL/6J, A/J, 129S1/SvlmJ, NOD/ShiLtJ, NZO/HiLtJ, and CAST/EiJ. For Experiment 2 mice from the two remaining progenitor strains (WSB/EiJ and PWK/PhJ) were obtained from Jackson Laboratories on a later date. C57BL/6J mice were also included in this group, as a standard comparison.

Experimental Design

Experiment 1

Five mice from each strain were exposed to a Chronic Ethanol Access (2) paradigm, and five were exposed to an Intermittent Ethanol Access (IEA) paradigm, each with three-bottle choice (15% and 30% ethanol v/v, and water), for 8 weeks. Ethanol mixtures were composed of anhydrous ethanol and tap water. Mice undergoing CEA had constant access to the three bottles, whereas mice undergoing IEA had 3 periods of 24-hour access (12 hours in the dark, then 12 hours in the light) each week (Monday, Wednesday and Friday), alternating with periods of access to three bottles of water only, as seen in Hwa et al. (36) (Fig. 1). Ethanol and water consumption were measured for each 24-hour period of access under IEA, for both paradigms, resulting in 3 recordings per week and 23 recordings total for each paradigm. On IEA ethanol access days, placement order of bottles was randomized for both paradigms to avoid place preference confounds. Bottles of each fluid were placed in uninhabited cages to measure evaporation and spillage. Measurements from these bottles were taken concurrently with consumption measurements, and were subtracted from that day's consumption values.

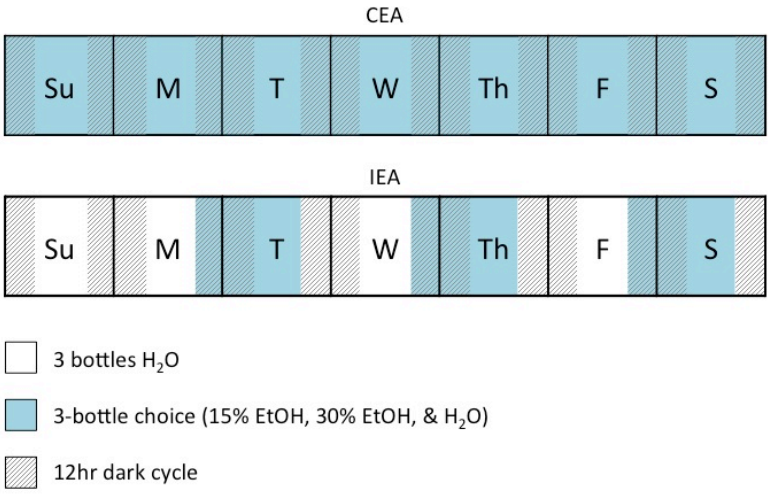


Figure 1: Experimental paradigms for Chronic Ethanol Access (2) and Intermittent Ethanol Access (IEA). Each block represents a day of the week, for which horizontal stripes represent the dark phases of the light-dark cycle, and solid areas represent light phases. White space represents periods during which mice had access to water only, and blue space represents period of ethanol access (1 bottle 15% EtOH, 1 bottle 30% EtOH, and 1 bottle water). These paradigms were repeated 8 times, resulting in 8 weeks of IEA or CEA.

On the 24th ethanol access day of IEA, ethanol bottles were removed for all mice to begin a 7-10 day period of forced abstinence. At the end of this period, mice received intraperitoneal injections of 4 g/kg of 20% v/v ethanol, for pharmacokinetic analysis. Trunk blood was collected during euthanasia via cervical dislocation and decapitation, which was staggered across the following time points: 30 mins, 60 mins, and 240 minutes. One to two mice per strain from each treatment group were euthanized at each time point. Blood Ethanol Concentration (BEC) was quantified via Analox AM1 Analyzer (Analox Instruments, North Yorkshire, UK), and recorded in mg_{EtOH}/dL_{blood} (or %BEC).

Experiment 2

Because there were no significant differences between paradigms in *Exp. 1*, all mice in *Exp. 2* were exposed to CEA, to increase power. Mice were exposed to 5 weeks of the same CEA paradigm described in *Exp. 1*. Additionally, to gather information about possible binge-like ethanol consumption in the PWK and WSB wild-derived strains, we measured ethanol and water consumption at 3 hrs after bottle placement, in addition to 24 hrs (as in

Exp. 1), and the additional time point of 3 hrs, after bottle order was changed for 5 consecutive days per week, to characterize the pace at which they consumed ethanol on the final day of CEA. After 3 weeks of CEA access, bottles were removed to begin a one-week period of forced abstinence.

As done in *Exp. 1*, following the week of forced abstinence, mice received intraperitoneal injections of 4 g/kg of 20% v/v ethanol. Trunk blood was collected during euthanasia via cervical dislocation and decapitation, which was staggered across the following time points, with 2 mice per time point: 15 mins, 30 mins, 60 mins, 120 mins, and 180 mins. These time points, which differ from those used in *Exp. 1*, were chosen to achieve a better estimation of the shape of the metabolism curve. Due to deaths by unknown causes during the abstinence period, data was collected from only one mouse for C57/BL6J at 15mins and PWK at 30mins. BEC was quantified using the same methods as *Exp. 1*.

Statistical Analysis

Experiment 1

Ethanol consumption was measured as g_{EtOH}/kg_{mouse}, to account for weight differences. Total ethanol consumption was calculated by summing the 15% and 30% ethanol consumption values. Total ethanol preference was calculated as $mL_{TotalEtOH}/mL_{TotalFluid}$ consumed. For heritability analysis and paradigm effect testing, each mouse's total ethanol consumption and preference were averaged across the entire study. Because these mice were ethanol-naïve on the first day of ethanol access, this day was excluded from these calculations. All statistical analyses were run using the "stats" and "DescTools" R packages (http://www.R-project.org/).

Heritability was calculated using one-way ANOVA, such that broad-sense heritability (h^2) = $SS_B/(SS_B+SS_W)$, where SS_B and SS_W represent the sum of squared error between and within strains, respectively ((149, 150)). Heritability was calculated for consumption and preference for each paradigm separately, and significance was determined by the *p*-value for mouse strain effect on the trait. Tukey Post-Hoc Test was performed to determine which strain differences contributed to significant heritability estimates.

Because we were interested in the distinct effect of paradigm on consumption and preference within each strain, we tested this effect via two-tailed Welch's t-test for each individual strain. We calculated False Discovery Rates to account for multiple testing within each trait.

To test for an increase in ethanol consumption or preference over time for each strain, every mouse's daily values were averaged over each the first and last week of ethanol access. We ran repeated-measures two-tailed Welch's T-tests for consumption and preference within each strain, under CEA and IEA separately. False Discovery Rates were calculated to account for multiple-testing within each paradigm for each trait. We were further interested in the difference in this effect across strains, also known as a gene x environment interaction (151). Where at least one strain showed significant or suggestive time (i.e. ethanol exposure duration) effects, we tested for gene x environment interaction (i.e. *time* x *strain* interaction) via two-way ANOVA. Although the main effects of strain and time period were not of interest here, they were included in each model to preclude false inflation of the interaction effect.

Pearson correlations were used to assess the relationship between mean total ethanol consumption and BEC, at each post-injection time point. Because there were < 3 mice per

strain at some time points, the contribution of metabolic variation to strain differences was assessed qualitatively. This was done by calculating the cumulative BEC across all post-injection time points and the ethanol elimination rate for each strain. Cumulative BEC was estimated by the area under the curve, based on strain mean BEC at each time point. Elimination rate was estimated by the slope of the same curve from point of highest measured BEC for that strain, to the final time point, calculated by linear regression. Because paradigm did not significantly effect ethanol consumption or preference (see *Results*), and all mice had been abstinent for ≥1 week, data were collapsed across treatment group for this analysis.

Experiment 2

Whole-study mean total ethanol consumption was calculated as seen in *Exp. 1*, for the 3hr and 24hr time points. The proportion of daily ethanol consumed during the first 3hrs of access was calculated for each day. These values were averaged across the entire study, again excluding the first day of ethanol access.

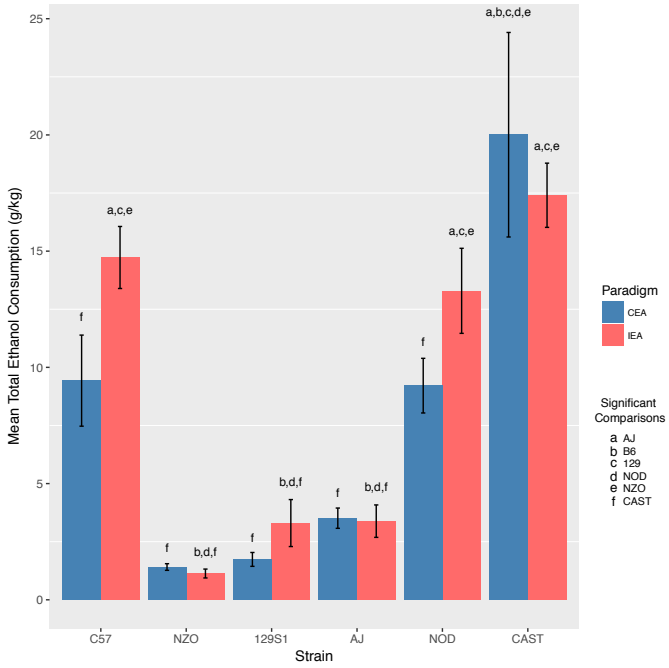
The same algorithms presented in *Exp.* 1 were used for the following analyses: heritability of mean total ethanol consumption and preference, and the associated post-hoc comparisons; change in consumption over time (first v. last week of ethanol access); and strain-specific ethanol kinetics and the relationship between BEC and total ethanol consumption. One-Way ANOVA was used to test for between-strain differences in proportion of daily ethanol consumed in the first 3hrs of access.

Results

Experiment 1

Paradigm did not have significant main effects on ethanol consumption or preference for any strain (Fig. 2; Table S2). There was also no significant increase in ethanol consumption or preference, for any strains after correction for multiple testing (Table S4a, Fig. 3a-b). However, increase in ethanol preference for NOD/ShiLtJ mice (t=2.41, p=0.022, q_{FDR} =0.199) was significant (p<0.05) before adjustment for multiple testing, under CEA. The same was true for total ethanol consumption under IEA in A/J mice, but the direction of change was negative (t=-2.84, p=0.042, q_{FDR} =0.190). Strain x time interactive effects on total ethanol consumption and preference, tested via two-way ANOVA, were not significant for either preference or consumption under CEA or IEA (Table S4b).

a



b

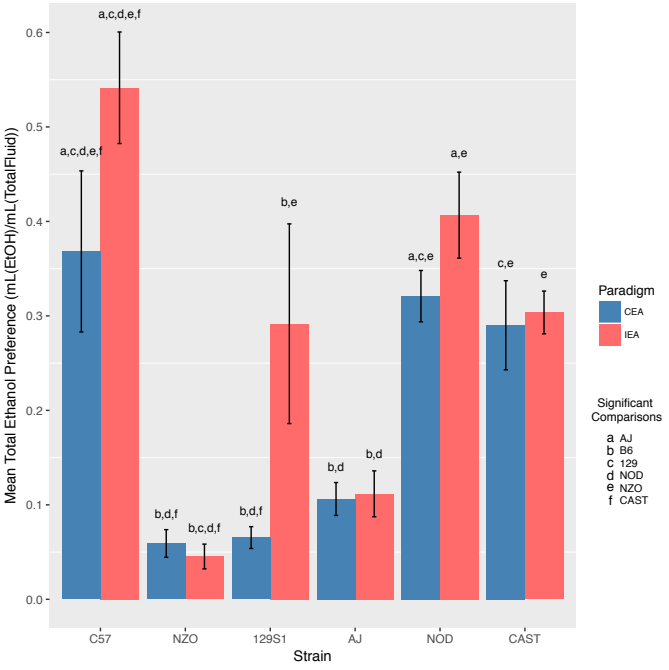


Figure 2: Average total ethanol consumption (a) and preference over water (b), across the entire experiment (except for access day1) for each strain in the first experimental group. Red and blue bars represent values under the Intermittent Ethanol Access (IEA) and Chronic Ethanol Access (2) paradigms, respectively. Black lines depict standard error around each mean, and letters represent significant differences between the strain depicted on x-axis and the corresponding strain listed in the legend.

There were large differences in ethanol consumption and preference across strains with highly significant heritability estimates under both CEA and IEA (Table 1). Post-hoc analysis revealed several significant pairwise comparisons across strains (Fig. 2, Table S1). For both phenotypes, the significant genetic effect stems from relatively low values for A/J, 129S1/SvlmJ, and NZO/HiLtJ compared to other strains, under each paradigm.

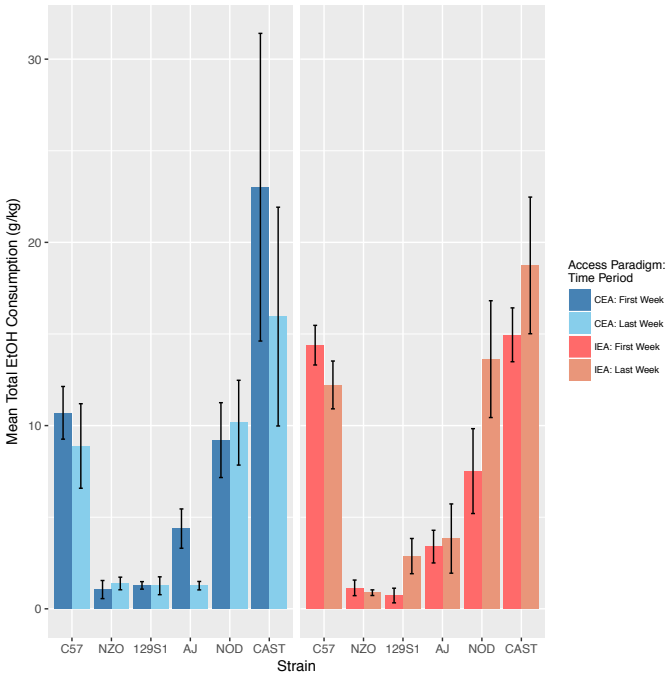
To assess whether consumption values might be modulated by differences in ethanol metabolism, we performed a pharmacokinetic analysis. Because there was no significant effect of paradigm on ethanol consumption or preference, and all mice had been abstinent for at least 1 week, mice were not separated by treatment group for this analysis. This analysis revealed a significant negative correlation between BEC and whole-study mean total ethanol consumption, 240mins after intraperitoneal ethanol injection (Fig. 4a). Although this correlation was not significant for other time points, there is a similar association pattern for initial BEC (30 mins) (Fig. 4a). Mice appear to cluster by strain, in accordance with this relationship. The cumulative BEC and ethanol elimination rate for each strain are represented in Table 2, and Fig. 4b.

Table 1: Heritability of Total Ethanol Consumption and Preference

Paradigm	Trait	Experiment	Strains	h^2	*p*
CEA	Consumption	1	C57BL/6J, 129S1, CAST, NOD, NZO, A/J	0.71	6.659E-06 ***
		2	C57BL/6J, WSB, PWK	0.78	1.695E-09 ***
	Preference	1	C57BL/6J, 129S1, CAST, NOD, NZO, A/J	0.69	1.606E-05 ***
		2	C57BL/6J, WSB, PWK	0.74	1.010E-08 ***
IEA	Consumption	1	C57BL/6J, 129S1, CAST, NOD, NZO, A/J	0.88	3.199E-10 ***
	Preference	1	C57BL/6J, 129S1, CAST, NOD, NZO, A/J	0.70	1.162E-05 ***

Heritability coefficients for total ethanol consumption and preference under chronic ethanol access (2) and intermittent ethanol access (IEA).
*** $p < 0.0001$

a



b

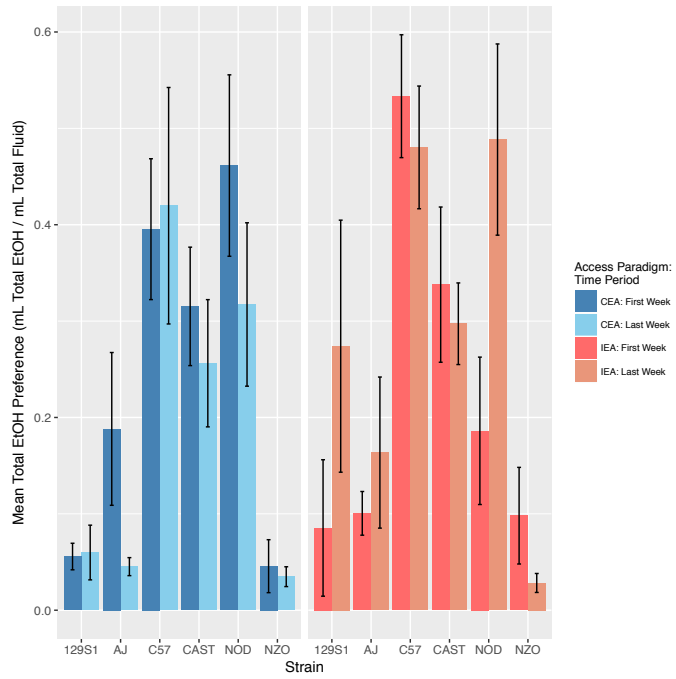
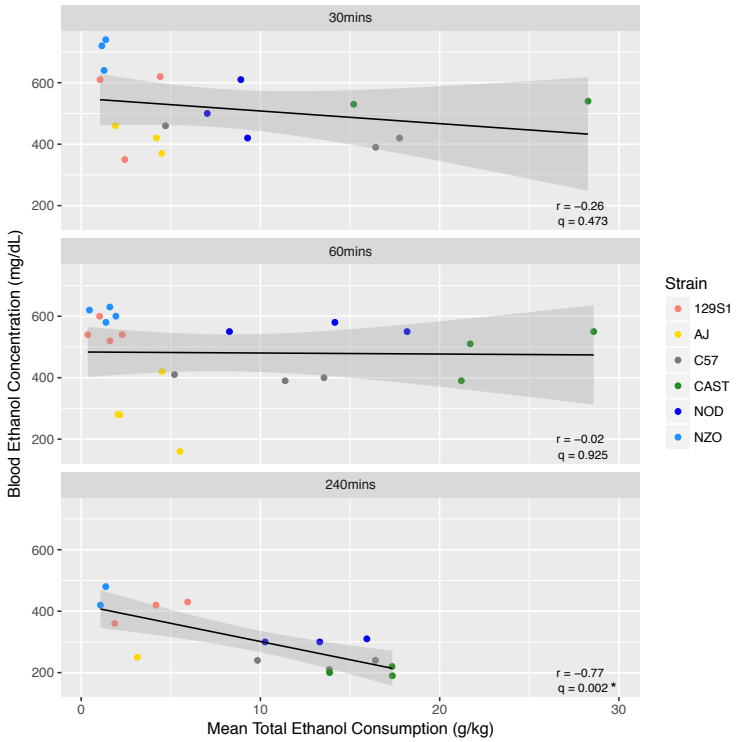
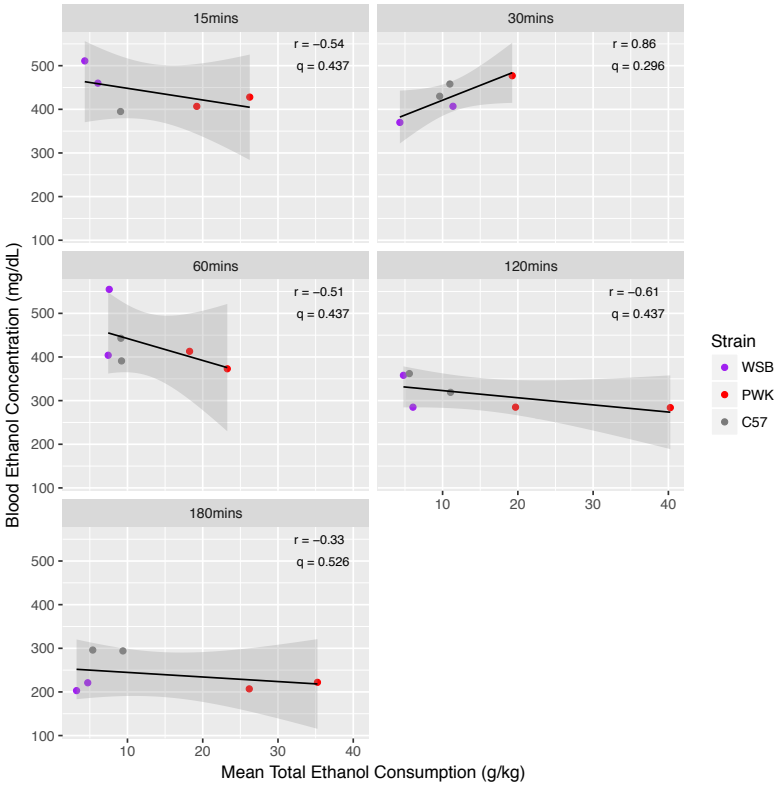


Figure 3 (a and b, see next slide for b plot) represents average total ethanol consumption (a) and preference for ethanol over water (b) for each strain, during the first and last weeks of ethanol access, with black lines representing standard errors. Dark blue and light blue bars represent values for the first and last week under Chronic Ethanol Access (2), respectively; and salmon and peach bars represent values for the first and last week under Intermittent Ethanol Access (IEA), respectively. There were no significant differences between time points for any strain.

a



b



c

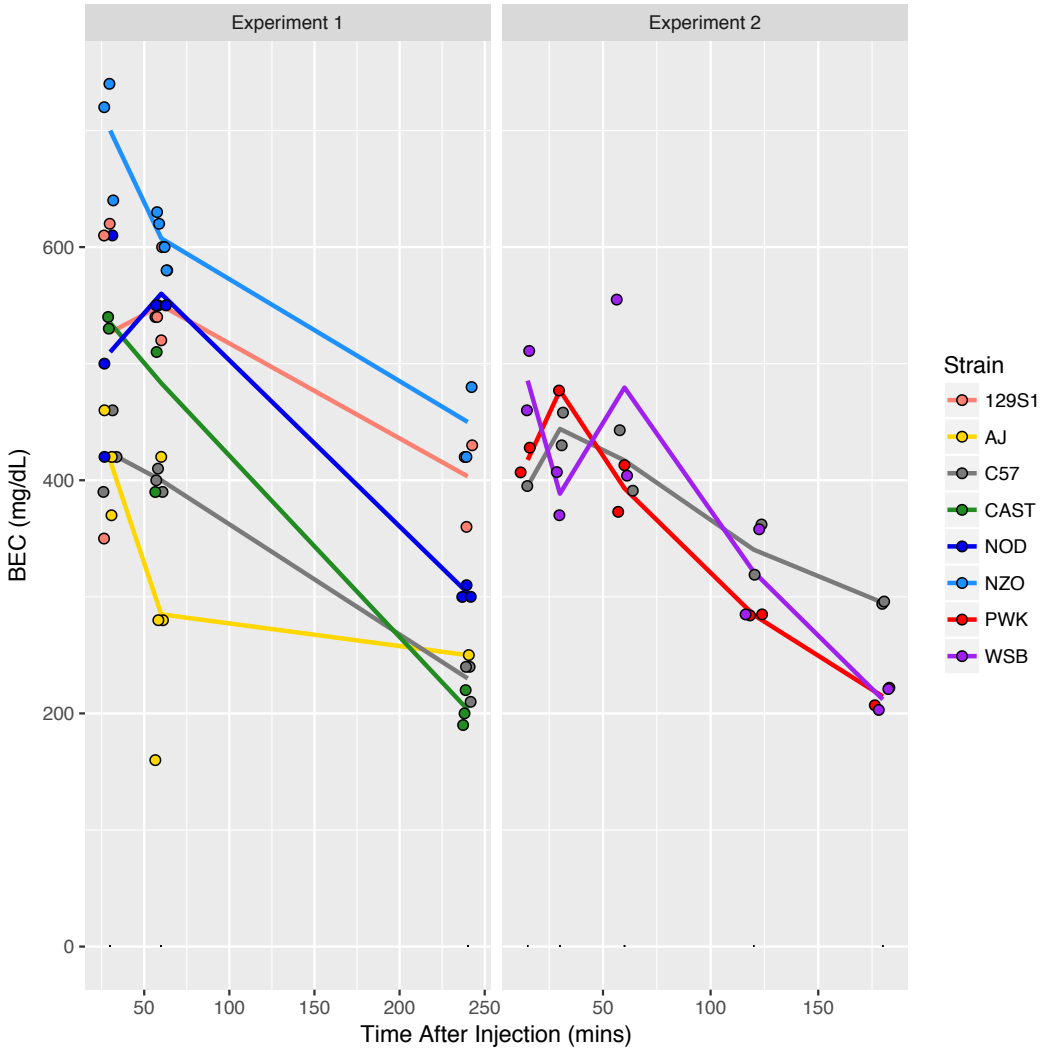


Figure 4: (a) blood ethanol concentrations (BEC) taken after intraperitoneal injections of 0.8g/kg(EtOH) at each of several time points following injection, depicted at the top of each plot, for Experiments 1 (left) and 2 (right). Each point represents one mouse, with color representing strain. These values are plotted over total ethanol consumption, averaged over the entire study, for each mouse. Black lines represent the line of best fit, and the surrounding gray-shaded area represents standard error. The corresponding correlation coefficient (r) and False Discovery Rate (q) are listed in the lower right corner of each plot, with * representing significance. (b) Blood ethanol concentration (BEC) values for each strain, for Experiments 1 (left) and 2 (right), across post-intraperitoneal-ethanol-injection time points. Each point represents one mouse, with colors representing strain. Lines run through the strain means at each time point, demonstrating the rate of change in BEC over time. Notches on x-axis represent time points at which BEC was measured.

Experiment 2

Just as in *Exp. 1*, there was no significant increase in ethanol consumption or preference over the course of the experiment, for any strains after correction for multiple testing (Table S4; Fig. 5a-b). However, there was a significant decrease in ethanol preference for WSB mice (t=-4.02, p=0.003, q_{FDR}=0.009) (Fig. 5b). Further, ANOVA results revealed a significant interaction between strain and time on ethanol preference only, meaning change in preference over time varies significantly between strains (Table 3; Fig. 5b). Notice that, in comparison with CEA consumption and preference values in *Exp. 1* (Fig. 3), C57 mice consumed ethanol at similar levels between experiments (although preference levels were slightly different), indicating reliability between experimental groups. However, PWK and WSB mice consumed and preferred ethanol at much higher levels than the other strains, meaning that they would have contributed to an even higher estimate of heritability, had the two experimental groups been combined.

Also similar to *Exp. 1*, both total ethanol consumption and preference were highly and significantly heritable (Table 1). Post-hoc analysis revealed significant differences between all three strains for total ethanol consumption, and between PWK and the other two strains for preference (Fig. 6a-b). However, the proportion of daily ethanol consumed in the first 3hrs of the dark phase did not differ significantly between strains (Fig. 6c). All three strains consumed roughly 14-15% of their daily ethanol during the first 3 hrs of access.

Although the correlation between BEC and whole-study mean total ethanol consumption was not significant, the relationship appears negative for all time points except 30 mins (Fig. 4a). As in *Exp. 1*, mice appear to cluster by strain, along the lines of best fit (Fig. 4b). The cumulative BEC and ethanol elimination rate for each strain are represented in Table 2.

Table 2: Blood Ethanol Concentration Statistics by Strain

Experiment	Strain	Slope From Peak	AUC	Time Point (mins)	N	Mean BEC mg/dL	BEC mg/dL SE	Max BEC
1	129S1	-0.81	101950.0	30	3	526.67	88.38	No
				60	4	550.00	17.32	Yes
				240	3	403.33	21.86	No
	AJ	-0.65	58675.0	30	3	416.67	26.03	Yes
				60	4	285.00	53.15	No
				240	1	250.00	NA	No
	C57	-0.93	69050.0	30	3	423.33	20.28	Yes
				60	3	400.00	5.77	No
				240	3	230.00	10.00	No
	CAST	-1.57	77075.0	30	2	535.00	5.00	Yes
				60	3	483.33	48.07	No
				240	3	203.33	8.82	No
	NOD	-1.43	93750.0	30	3	510.00	55.08	No
				60	3	560.00	10.00	Yes
				240	3	303.33	3.33	No
	NZO	-1.08	114787.5	30	3	700.00	30.55	Yes
				60	4	607.50	11.09	No
				240	2	450.00	30.00	No
2	WSB	-1.52	59610.0	15	2	485.50	25.50	Yes
				30	2	388.50	18.50	No
				60	2	479.50	75.50	No
				120	2	321.50	36.50	No
				180	2	212.00	9.00	No
	PWK	-1.65	55053.0	15	2	417.40	10.60	No
				30	1	477.00	NA	Yes
				60	2	393.00	20.00	No
				120	2	284.50	0.50	No
				180	2	214.50	7.50	No
	C57	-1.03	60997.5	15	1	395.00	NA	No
				30	2	444.00	14.00	Yes
				60	2	417.00	26.00	No
				120	2	340.50	21.50	No
				180	2	295.00	1.00	No

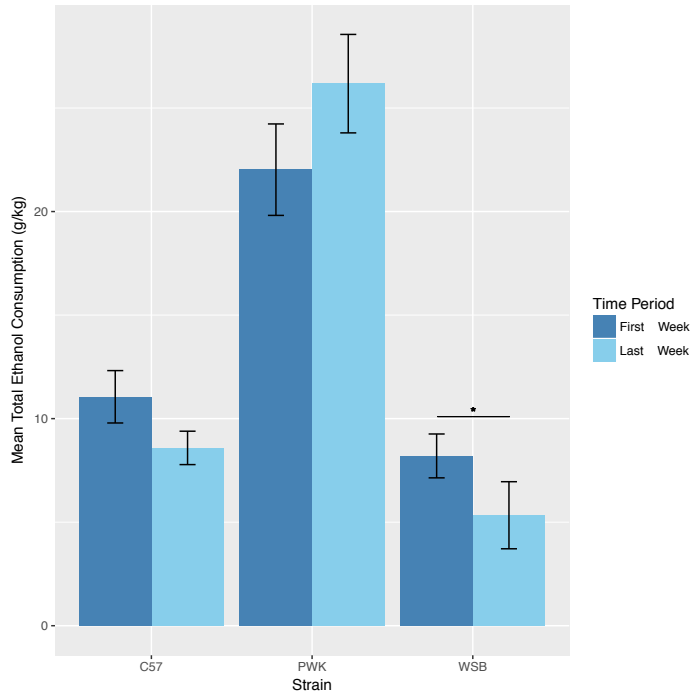
Slope, area under the curve (AUC), timespan between ethanol treatment and blood collection, number of mice in each group (N), mean and standard deviation (SD) blood ethanol concentration (BEC), and whether or not the maximum BEC for the respective strain occurred the respective time point.

Table 3: ANOVA Results for Effect of Time and Strain on Drinking Behaviors

Trait	I.V.	df	F value	p	q_{FDR}
Consumption	Strain	2	61.09	1.740E-14 ***	5.221E-14 ***
	TimePeriod	1	0.07	7.966E-01	7.966E-01
	Strain:TimePeriod	2	2.77	7.162E-02	1.074E-01
Preference	Strain	2	47.96	1.289E-12 ***	3.868E-12 ***
	TimePeriod	1	9.75	2.904E-03 **	4.355E-03 **
	Strain:TimePeriod	2	4.75	1.271E-02 *	1.271E-02 *

The F-statistic, degrees of freedom (1), p-value, and FDR-adjusted q-value for ANOVAs testing for the effects of each independent variable (IV) on total ethanol consumption and preference.
* $p<0.05$; ** $p<0.01$; *** $p<0.001$

a



b

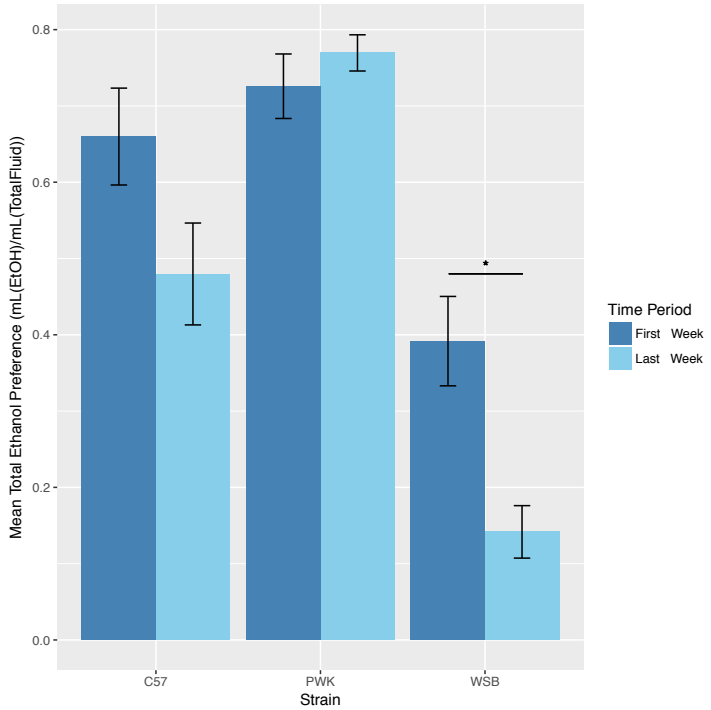


Figure 5: Mean total ethanol consumption (a) and preference over water (b) for the first (dark blue) and last (light blue) weeks of Chronic Ethanol Access, for each strain in Experiment 2. Bars with asterisks represent significant differences between time points for that strain.

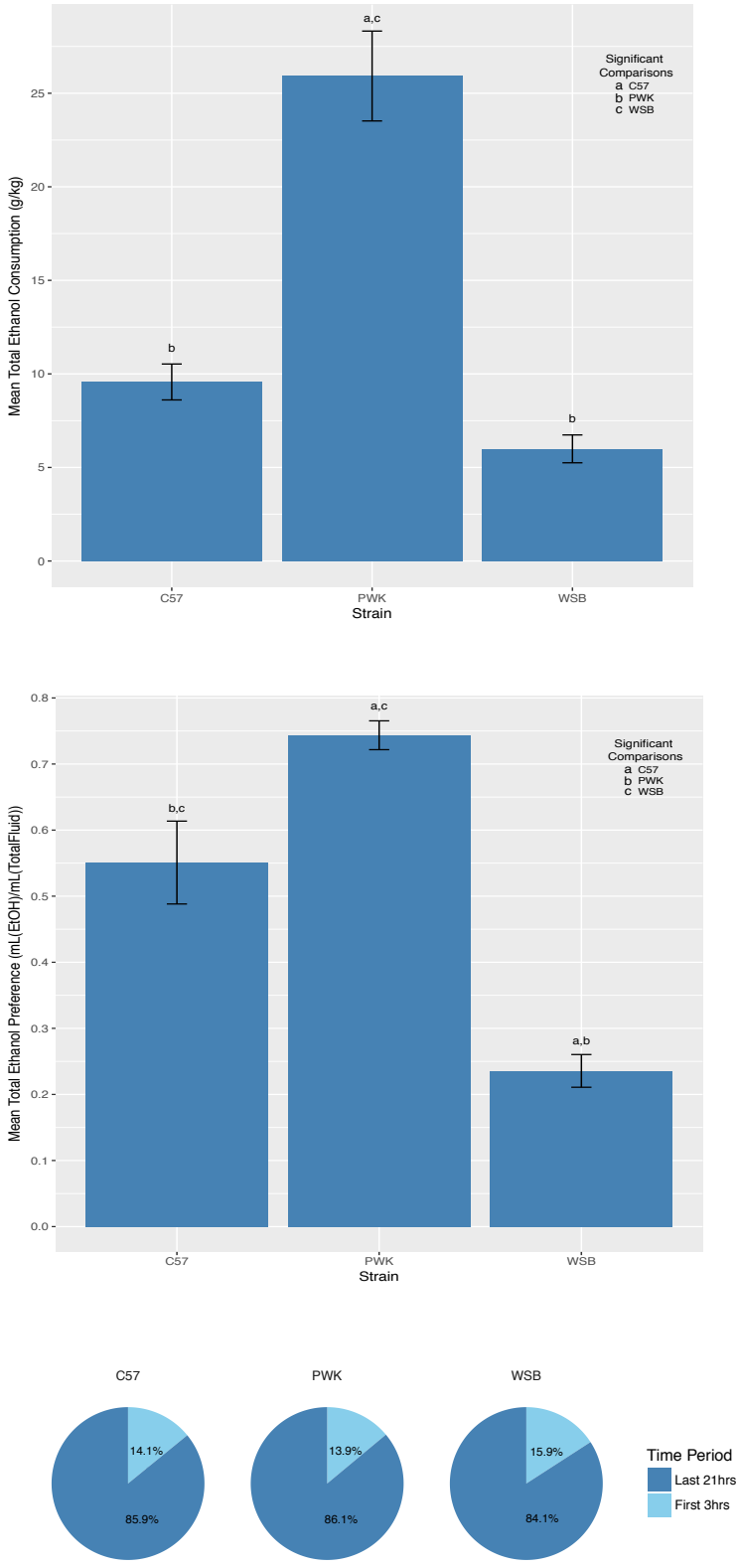


Figure 6: Average total ethanol consumption (a) and preference (b) under Chronic Ethanol Access for each strain in Experiment 2, across the entire study. Black lines depict standard error around each mean, and letters represent significant differences between the strain depicted on x-axis and the corresponding strain listed in the legend. (c) Average proportion of daily total ethanol consumed in the first 3hrs of ethanol access (light blue) compared to the remaining 21hrs of access (dark blue), for each strain. There were no significant differences in proportions seen between strains.

Discussion

The goal of this study was to characterize voluntary ethanol consumption levels and ethanol preference in each of the Diversity Outbred stock progenitor strains, and to determine the feasibility of genetically mapping these traits in a DO sample. We accomplished this by longitudinally measuring voluntary ethanol and water consumption in mice from each of the progenitor strains during exposure to several weeks of Intermittent or Chronic Ethanol Access paradigms, in the absence of prior ethanol treatment or exposure. Specifically, we determined the heritability of total ethanol consumption and preference over water, and characterized within-strain consistency of these characteristics across paradigms and experimental time points. We also measured BEC in these mice at various time points following intraperitoneal ethanol injection to examine the potential contribution of metabolic variation to strain differences in ethanol consumption and preference. Overall, our studies show highly significant and heritable cross-strain variation in ethanol consumption and preference and suggestive strain differences in the temporal pattern of ethanol consumption, supporting use of the DO mice for genetic mapping of ethanol drinking behaviors.

IEA v. CEA

Within-strain analysis suggested that total ethanol consumption and preference behaviors were consistent across paradigms in *Exp. 1* (examining C57, AJ, NOD, NZO, 129S1, and CAST), resulting in the use of CEA only in *Exp. 2* (examining C57, PWK, and WSB). These phenotypes were also consistent across time for most strains, under both the intermittent and chronic paradigms. However, there was a significant decrease in ethanol preference for WSB mice in *Exp. 2* (which only examined CEA), and a trending decrease in ethanol consumption for A/J mice, under IEA only. There was also a trending (significant

before multiple testing adjustment) increase in ethanol preference for NOD/ShiLtJ mice, under CEA only. Despite these differences, there was no significant time x strain interaction in either experiment for ethanol consumption, or for ethanol preference in *Exp 1*. This interaction was significant for ethanol preference in *Exp 2*. However, instead of having positive slopes of different grades, C57 and WSB mice actually decreased their preference over time, while PWK increased. These findings are inconsistent with findings from previous studies on C57BL/6J mice , which suggest that intermittent access leads to greater overall ethanol consumption, and a greater increase in ethanol consumption and preference compared to chronic access in C57BL/6J mice.

One possible explanation for the similarities in phenotypes between IEA and CEA in this study is that we employed an IEA paradigm with 24hr abstinence periods, which likely did not allow for a true alcohol deprivation effect (148) to occur. Another possible contributor is the use of a 3-bottle choice model, as opposed to 2-bottle choice, which likely increased within-group variation. Finally, although we used the IEA paradigm reported in Hwa et al., we did not implement an initial period of habituation to progressively increasing concentrations of ethanol, due to the use of a 3-bottle choice paradigm. The absence of this habituation period may have created a ceiling effect (with ceiling levels varying between strains), meaning the mice were not driven to drink over a certain amount, regardless of access frequency. Also counter to what we hypothesized, the effect of intermittent v. chronic access on ethanol consumption (gene x environment interaction) did not differ between strains. Although Rosenwasser et al. (79) saw higher ethanol consumption under IEA versus chronic access in only 4 (C57BL/6J, C3H/HeJ, HDID-1, and HS/Npt) of the 6 (WSP-1 and WSR-2) strains they studied, they studied different strains from the present study (aside from

C57BL/6J), and employed an IEA paradigm with longer periods of abstinence. Additionally, they did not test for a *paradigm* x *strain* interaction directly. To our knowledge, this study is the first to directly test the effect of this interaction and the *paradigm* x *strain* interaction on ethanol consumption and preference.

Heritability

Our between-strain analysis showed moderate to high heritability of voluntary ethanol consumption and preference, under intermittent and chronic access paradigms among the DO progenitor strains. Post-hoc analysis indicated that 129S1, NZO, and A/J mice will not voluntarily consume ethanol at high levels, and do not prefer it over water, unlike the remaining DO progenitor strains, particularly the PWK and CAST wild-derived strains. These findings are consistent with those of previous studies, which have shown much lower consumption and preference values for A/J and 129S1 mice, and slightly lower values for NOD mice, than C57BL/6J mice (78, 152, 153). These results suggest that this trait possesses the variance between strains necessary for detecting quantitative trait loci in an outbred mouse panel derived from these strains. Although results from such studies would likely be highly complex and polygenic in nature, they would likely identify some loci within haplotypes inherited from 129S1, NZO, or AJ progenitors, with protective effects (i.e. confer lower consumption levels or preference), and some loci within haplotypes inherited from PWK or CAST, conferring risk for higher consumption levels or preference. Moreover, these results demonstrate the uniqueness of wild-derived strains, supporting the theory (32, 90) that they possess behavioral traits and contributing loci that could not be detected in studies of only lab-derived strains.

Although PWK mice had higher ethanol consumption than C57 and WSB, in *Exp. 2*, the amount of ethanol they consumed in the first 3 hrs of ethanol access did not represent a higher proportion of their daily consumption than the other two strains. This indicates that they all steadily consumed ethanol over a 24 hr period (3hrs/24hrs=12%), as opposed to binging. In other words, the PWK mice consistently drank more ethanol over the course of the 24 hrs of access than the other two strains, as opposed to acutely consuming large amounts of ethanol at the onset of ethanol access, then "sobering up" for the duration of that access period. These results suggest that PWK mice may metabolize ethanol differently than the other two strains, as they are able to sustain high levels of consumption over a 24-hr period.

Ethanol pharmacokinetics

Overall relative relationship of ethanol elimination rates between strains (Table 2) were consistent with findings from Crabbe et al. (154), which examined ethanol elimination rate following IP injections of 2g/kg EtOH in C57, 129S1, A/J, NOD, and CAST, among other strains. In both studies, CAST had the fastest elimination rate, and A/J and 129S1 had the slowest rates, with C57 falling in between. However, in Crabbe et al., NOD had the same rate as AJ, whereas NOD's rate was faster than C57 in the present study. This could be due to the use of different BEC measurement time points between studies. The final time point in Crabbe et al. was 150mins, whereas our final time point in *Exp. 1* was 240mins to our knowledge no prior studies have characterized ethanol kinetics in PWK, WSB, or NZO.

Our preliminary BEC analyses suggest that the heritability of ethanol consumption and preference between the DO progenitor strains is likely partly attributable to differences in ethanol pharmacokinetics. There was a significant correlation between BEC and mean

total ethanol consumption at the 240min time point after intraperitoneal ethanol injections in *Exp. 1*. As would be expected, the lower the BEC, the more ethanol the mice drank, suggesting that mice that consume higher amounts of ethanol have a faster ethanol metabolism. Unexpectedly, this relationship was not significant for earlier time points in *Exps. 1 or 2*. Moreover, mean total ethanol consumption within strains appears to be independent of strain-wise cumulative BEC (area under the curve) and ethanol elimination rate (slope).

Together, these findings indicate that overall ethanol kinetics plays a role in the heritability of voluntary ethanol consumption. However, this relationship seems to be driven by BEC levels after several hours of metabolism, as opposed to the initial ethanol concentrations that might reflect differences in volume of distribution. These results are also consistent with findings from Crabbe et al. (154), which found that BEC 35mins after IP injection of 2g/kg EtOH did not correlate with ethanol sensitivity.

Shortcomings and Future Directions

As mentioned above, the use of 3-bottle choice and the absence of an ethanol habituation period may have precluded the detection of differences between intermittent and chronic paradigms, due to increased within-strain variance. Moreover, the size of this pilot study may have limited our power to detect differences across paradigms or time points.

Overall, this study exhibits that these progenitor strains offer a large degree of inter-strain variability in ethanol consumption, preference, and pharmacokinetics. Future Quantitative Trait Locus studies on Diversity Outbred mice will likely be able to map genetic loci contributing to voluntary ethanol consumption and preference. Markers within ethanol metabolism genes are likely to be amongst the significantly associated loci. These studies

also characterized longitudinal voluntary ethanol consumption and preference under chronic and intermittent paradigms, within each of the DO progenitor strains, and suggested possible strain-dependent temporal patterns that might also be amenable to genetic dissection in DO mice. Such studies could have high relevance to identifying genetic mechanisms contributing to the progression from social to abusive ethanol consumption.

Chapter 4

Mapping Precise Quantitative Trait Loci for Voluntary Ethanol Consumption in Diversity Outbred Mice

Introduction

Alcohol Use Disorder (AUD) is a debilitating illness that has many negative health and behavioral outcomes, with even short-term abstinence being as low as 43% for those who seek pharmacological or cognitive-behavioral treatment (11). One way to gain a better understanding of the biological processes, and therefore improve treatment, is to explore genetic contributions to the disorder. However, despite the 50% heritability estimated by twin studies (21), Genome-Wide Association Studies have had limited success in identifying significant and replicable loci. Unlike human studies, environmental factors that may muddle genetic signals can be controlled in mouse experiments. Although mouse models do not account for all of the nuances of human behavior, they are useful for identifying broadly applicable biological mechanisms underlying specific behaviors associated with a disorder, such as alcohol consumption and preference. In fact, mouse models have successfully identified several genetic loci associated with these two traits (65). However, these studies are often plagued by limited genetic heterogeneity (due to a limited number of progenitor strains) and low precision due to large linkage disequilibrium blocks, or regions of co-inherited DNA with limited amounts of detectable genetic cross-over (due to a small number of cross-breeding generations) (57).

To address these issues, this study aims to identify genetic loci associated with alcohol consumption and preference in a genetically heterogeneous mouse population from the Diversity Outbred (DO) stock. This stock was especially created for the precise mapping of genetic loci associated with complex behavioral traits, by interbreeding across Collaborative Cross recombinant lines, followed by several generations of random outbreeding (155). The Collaborative Cross mice were created from 144 different breeding

patterns, each incorporating 8 progenitor strains, 3 of which are wild-derived (155). These 8 strains afford the DO mice with roughly 4.5 times as many variants as there are in the human genome, and 36 possible diplotypes, which can be inferred from these SNPs, at each locus across the genome (156). This genetic diversity and complex breeding scheme make for a mouse stock with wide behavioral variability (similar to that seen within a human population), a vast array of detectable polymorphisms (including those present in the genome of wild mice), and smaller discernable haplotypic blocks than most other laboratory mouse strains. Despite their utility, to our knowledge, alcohol drinking phenotypes have not yet been genetically mapped using the DO mice.

In particular, this study aims to map behavioral Quantitative Trait Loci (bQTL), or regions of DNA within which genetic variants are associated with a behavioral trait, for consumption and preference under an Intermittent Ethanol Access (IEA) paradigm. This paradigm models relapse-like drinking seen in the early stages of dependence development with alternating periods of ethanol access and forced abstinence, which leads to increased ethanol consumption over time in some mouse strains (36). Ethanol consumption and preference have been determined to be highly heritable under this paradigm across the 8 DO progenitor strains, indicating that they can be genetically mapped in DO mice, as seen in Chapter 3. This experiment is anticipated to identify previously published as well as novel loci associated with ethanol consumption and preference, with greater precision than any previous study. Specific to this paradigm, loci associated with naïve ethanol consumption and preference at the beginning of the study are expected to differ from those associated with consumption after repeated exposure to abstinence. These findings would provide insight to

biological mechanisms involved in the development of progressive ethanol consumption that could potentially be targeted for treatment of AUD.

Materials and Methods

Ethics Statement

All animal care and euthanasia procedures were performed in accordance with the rules and regulations established by the United States Department of Agriculture Animal Welfare Act and Regulations, Public Health Services Policy on Humane Care and Use of Laboratory Animals, and American Association for Accreditation of Laboratory Animal Care**. Humane endpoints were established by the same standards.

Behavioral Experimentation

Male DO mice (N=636) were acquired from Jackson Laboratories at 4 weeks of age, in cohorts of 106 mice on average, spanning across DO generations 21-24. Mice were singly housed on cedar shaving bedding, with ad libitum access to water and standard chow (#7912, Harlan Teklad, Madison, WI, United States), and alternating 12hr light and dark phases, upon arrival and remained housed under these conditions for the remainder of the study. All mice were weighed weekly. Given their likely genetically driven ethanol-related behavioral differences, females were excluded from this study, so as to avoid losing power due to sex effects (157-159).

At between 8 and 12 weeks of age, adult mice underwent a 10min paradigm of anxiety-like behavioral testing in light-dark boxes. Mice were placed in Med Associates Inc. Mouse Locomotor Boxes with light-dark inserts to create a “dark” side and a “light” side.

Laser beams were used to track mouse movement between sides of the box, with longer time spent in the light considered to implicated lower levels of anxiety-like behavior. Mice were habituated in the experimental room that houses these boxes for 1hr prior to testing. Lateral and vertical movements were recorded across 5min time intervals, during the 10min exposure. Measurements calculated included, but were not limited to: time spent in the light, percent time spent in the light, total distance traveled, distance traveled in the light, percent distance traveled in the light, and stereotypy. This data was collected for future analyses on the relationship between drinking levels and anxiety-like behavior before and after (see marble burying test described below) ethanol exposure, and will not be discussed further in this manuscript. The following day, the Intermittent Ethanol Access (IEA) paradigm was initiated. This paradigm was consistent with that described in Chapter 3 with alternating 24hr periods of water and ethanol (Mon. Wed., Fri.), and water only (Sat., Sun., Tues., Thurs.) (Fig. 1). For ethanol exposure periods, three bottles (15% and 30% ethanol, and water) were placed in cages at the beginning of the dark phase, and consumption was measured in mL of fluid at the end of the following light phase, at which point ethanol bottles were replaced with water bottles. Controls (N=49) were exposed to water only, with water consumption measured on the treatment group's ethanol access days, throughout the course of the experiment.

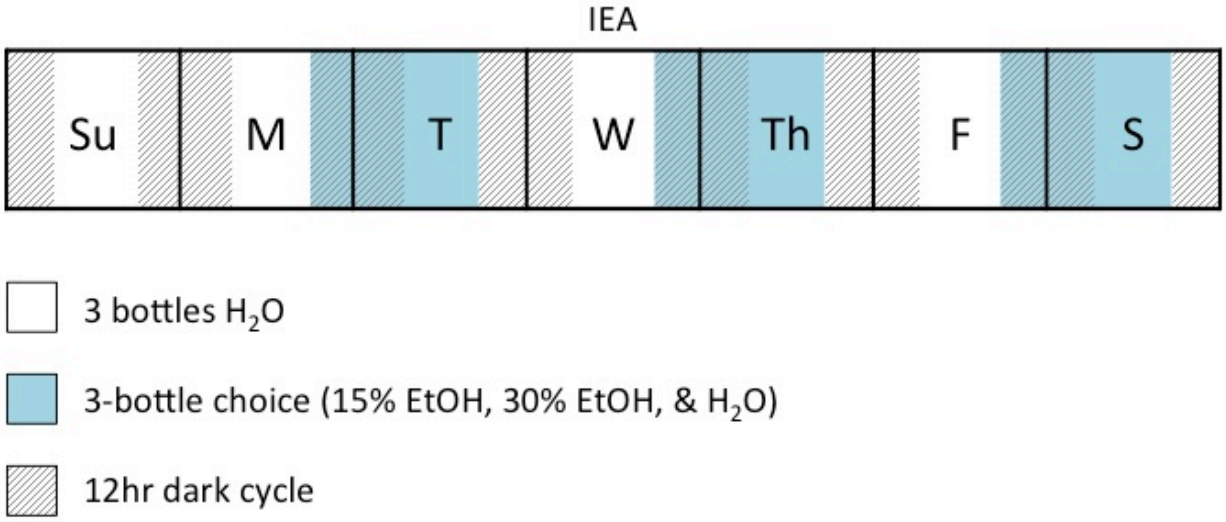


Figure 1: Intermittent Ethanol Access (IEA) paradigm employed in this study. Each block represents a day of the week, with dashed area indicating the 12hr dark phase of each 24hr light-dark cycle. Blue coloration represents 24hr periods of access to ethanol, with 3-bottle choice (15% EtOH, 30% EtOH, and water), and uncolored areas represent periods of forced abstinence, with access to 3 bottles of water only. This was repeated for 4 weeks prior to anxiety-like behavioral testing, and 1 additional week to reestablish unperturbed drinking behaviors and gene expression levels.

After 4 weeks of IEA, another anxiety-like behavioral test, marble burying, was performed. For this test, mice were placed in 13" x 15" rat cages filled 3/4 full with cedar chip bedding, and 16 marbles placed on top of the bedding. Mice were left in these boxes for a total of 20mins, and the number of marbles at least 2/3 buried was counted. These results can be interpreted as a larger number of marbles being buried indicating anxiety-like behavior. Like the light-dark box data, this data was collected for ancillary analysis, and will not be discussed further. Mice were then exposed to an additional week of IEA to re-establish gene expression patterns that may have been disrupted by other behavioral tests, for future gene expression-related experiments. Only the first 4, uninterrupted weeks of drinking behaviors were used for the purposes of this study. Mice were sacrificed via cervical dislocation and decapitation, 24hrs after the end of their last ethanol exposure period, and tissue samples were immediately collected and dissected for genotyping and gene expression assays. Tail snips, specifically, were collected for genotyping. Liver, spleen, and brain tissue were also collected for gene expression analyses performed in Chapter 5, and for future studies. Brains were microdissected into 9 regions, including the Prefrontal Cortex. Tissue was collected from 630 mice, as 6 mice reached human endpoints due to likely congenital health problems, before the end of experimentation. All tissue was immediately flash frozen after dissection, to preserve RNA integrity.

Ethanol consumption was measured as g(EtOH)/kg(mouse), to control for effects of mouse size on fluid consumption. Daily consumption values were marked as missing when: any one bottle had been emptied past the sipper opening; the bedding under the bottles was wet, indicating spillage or leakage; the sippers had been tampered with (ex. woodchip bedding packed into the opening) in a way that would prevent drinking or cause leaking; or

the consumption value were large negatives, indicating irreconcilable recording error. Values were marked as 0mL if the readings yielded a slight negative value (down to roughly -0.3mL), indicating inter-rater variation on a true 0 read. Because of the expectation of wide variation in such a genetically diverse sample, only extreme outlying values were removed, leading to one sample's consumption values being removed for one access day. If more than half of the values needed for weekly or whole-study mean calculations were missing for any given mouse (due to bottle leakage or complete emptying of any one of the 3 bottles), that sample's mean was marked as missing. Total ethanol consumption was calculated by summing the g/kg values for 15% and 30% ethanol. Total ethanol preference was represented by the fraction of mL total ethanol (mL(15%EtOH) + mL(30%EtOH)) over mL total fluid (mL(water + mL(15%EtOH) + mL(30%EtOH)) intake. Finally, 30% ethanol preference over 15% ethanol, hereafter referred to as "30% choice", was represented by the fraction of g/kg 30% ethanol over g/kg total ethanol consumed. For QTL mapping, all values were averaged over the first week (the 3 days following the first day), the last week (the final 3 days of ethanol access), and the whole study course (all 10 ethanol access days, excluding the first day) of IEA. First and last week values are meant to map genetic loci associated with naïve and post-exposure drinking, respectively; whereas whole study values are meant to capture loci associated with overall drinking behaviors, irrespective of exposure time.

Genotyping

Tail tissue from each mouse was sent to NeoGen Inc. (Lincoln, NE) for DNA isolation and microarray genotyping. The GigaMUGA microarray, created specifically for identifying single nucleotide polymorphisms (SNPs; N_{SNP}=141,090) and copy number

variants (CNVs and indels; N_{CNV}=2,169) that are informative between the DO mouse stock's 8 progenitor strains (160). Mice were genotyped in two batches (N_{Batch1}=210, N_{Batch2}=420). Four samples from the first batch were run again on the sample batch, due to low quality. Three high-quality samples from the first batch were run with the second batch, so that the results between the two arrays could be compared, in order to detect batch effects. Quality and genotype calls were consistent between arrays, indicating little technical variation (Fig. 2-5). (*Genotyping batch* correlates perfectly with *Cohort*, so the inclusion of *Cohort* as a covariate in our analyses corrected for any undetected differences between *genotyping batches*.)

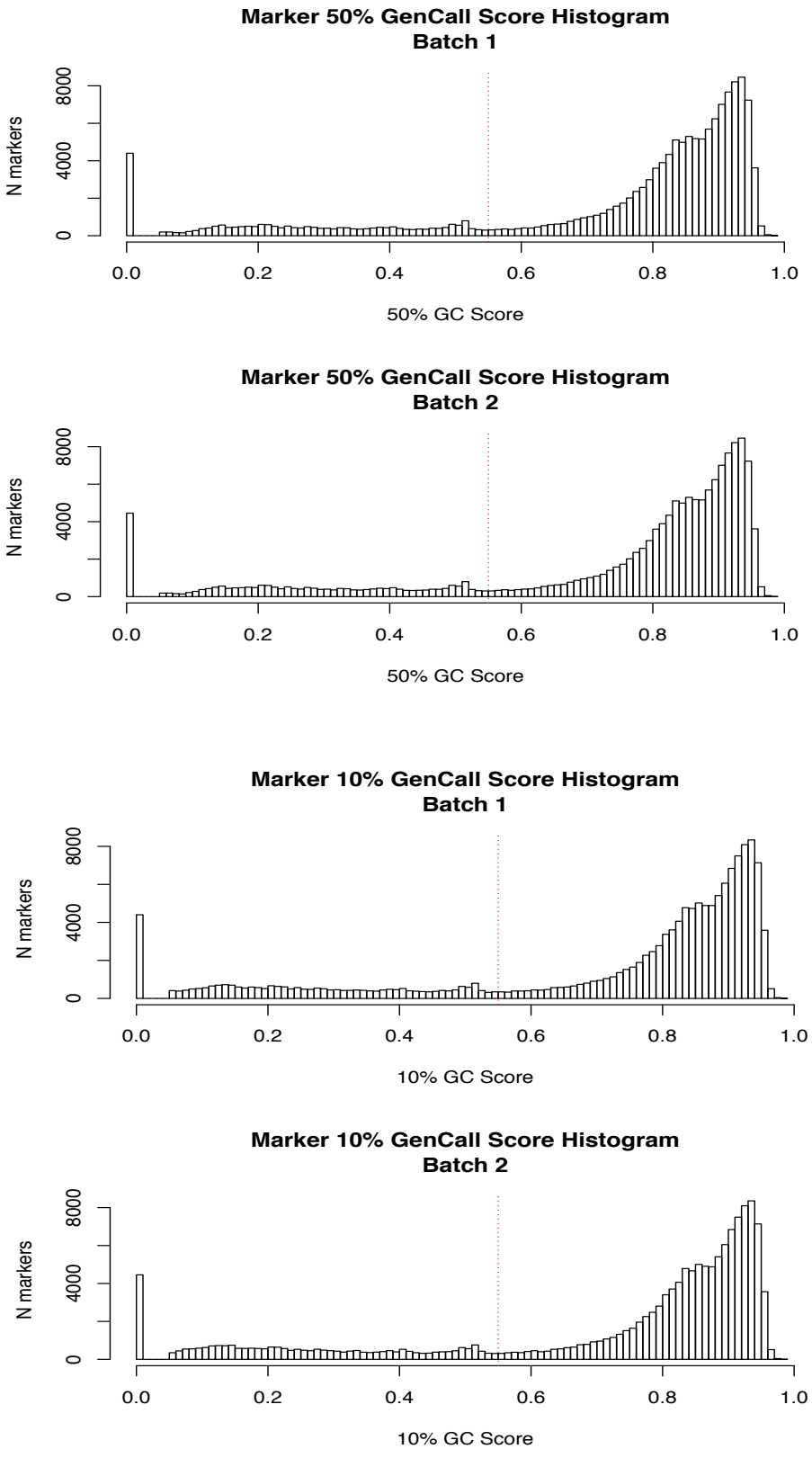


Figure 2: Histograms of 50% (left) and 10% (right) Marker GenCall scores for genotyping batches 1 (top) and 2 (bottom). Dotted red lines represent the threshold scores. Markers with 10% scores that fell below this line in both batches, or with 50% scores that fell below this line in either batch, were removed during quality control. Scores were similarly distributed between batches.

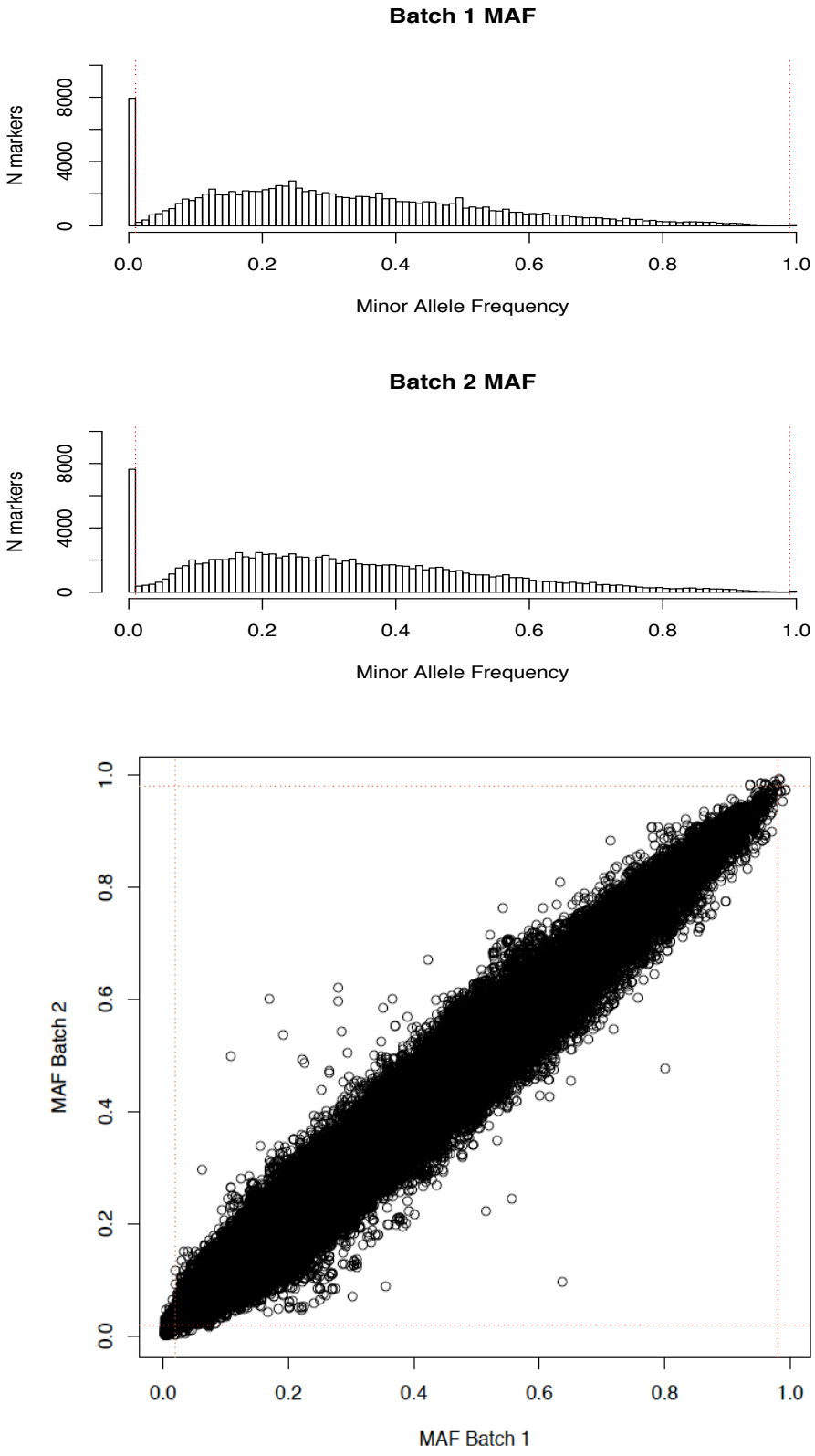


Figure 3: Minor allele frequencies (MAFs) for each genotyping batch. On the top are histograms of MAFs in each batch, with red dotted lines representing cutoff thresholds for quality control. Markers between these lines were kept, and markers that fell outside of these lines were removed. On the bottom is the relationship of MAFs between batches, with quality thresholds marked by dotted red lines. MAF was highly similar between batches.

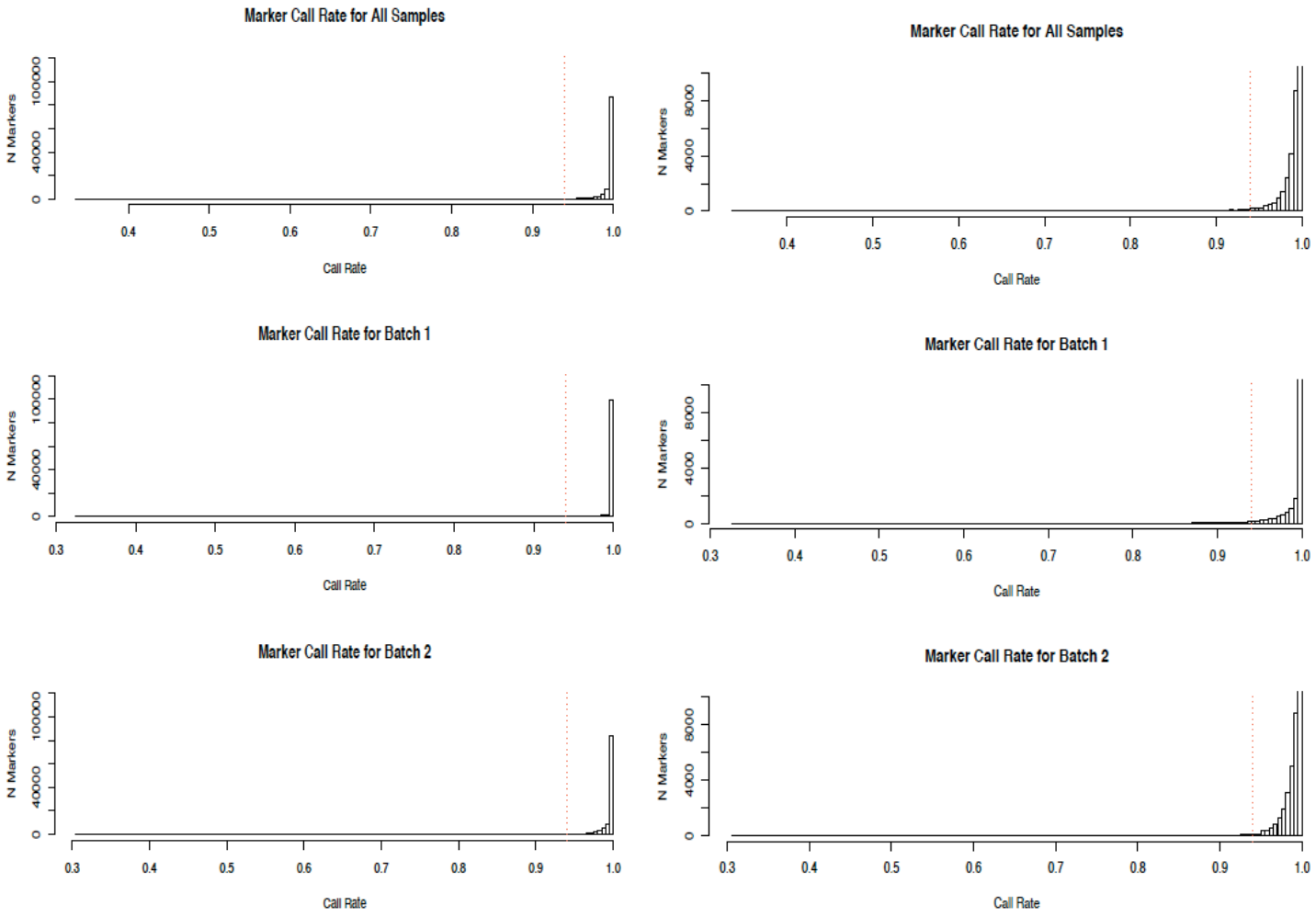


Figure 4: Marker call rate histograms (left) across all samples (top), then broken down by batch (middle and bottom). The right-hand plots are Y-axis-truncated versions of the left-hand plots, so that the shape of the distributions can be more easily seen. Red dotted lines indicate quality thresholds, below which markers were removed.

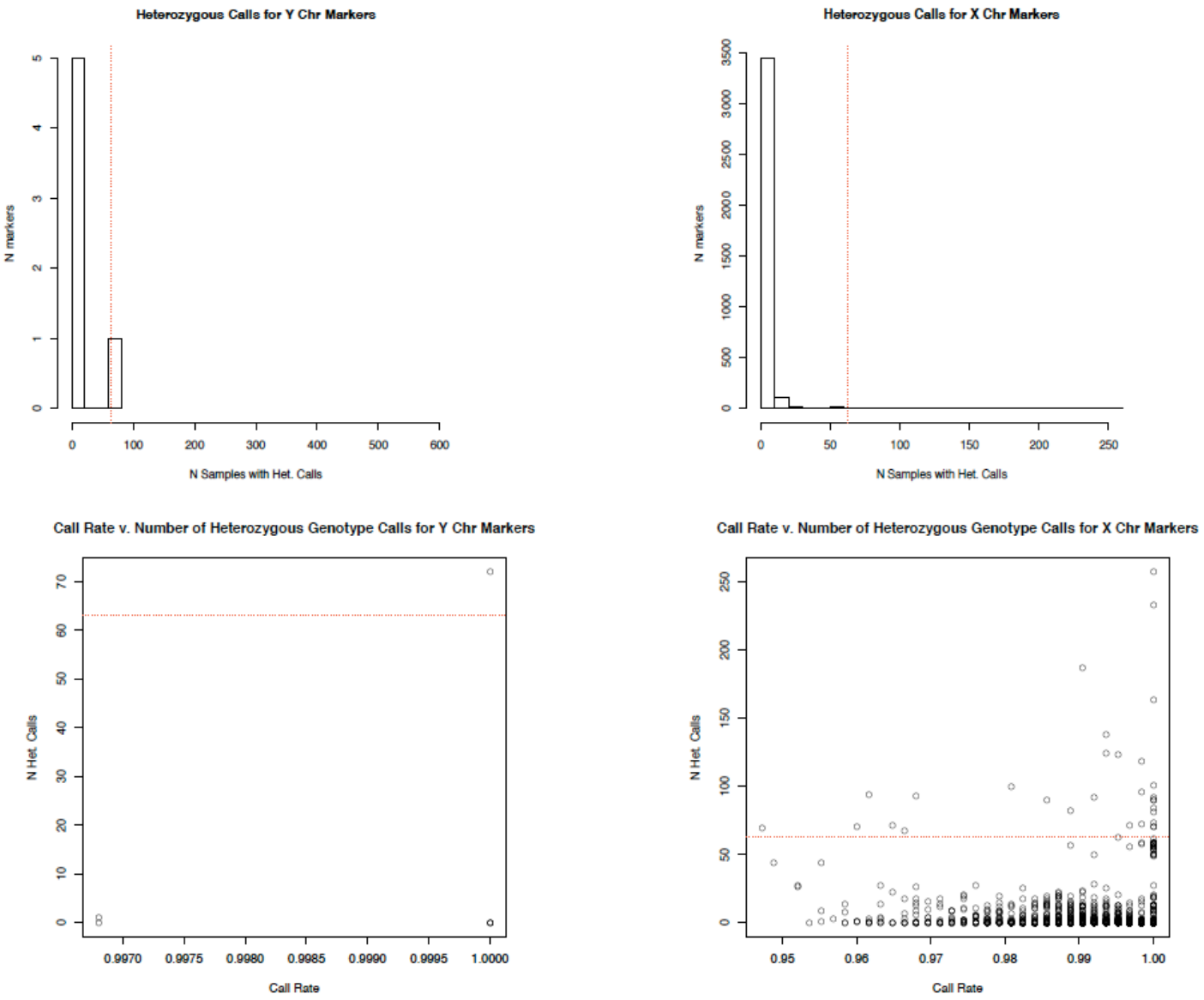


Figure 5: Histograms of heterozygous calls for the Y (left) and X (right) chromosomes (top), with red dotted lines indicating quality thresholds above which samples were removed. Bottom plots depict overall sample-wise call rates by the number of heterozygous calls for the Y (left) and X (right) chromosomes, with red dotted lines indicating the same thresholds depicted in the upper plots. In general, markers with higher numbers of heterozygous calls also appear to have larger call rates; so using only a call rate threshold would not have been sufficient quality control for X and Y chromosome markers.

Marker-Level Quality Control

All quality control steps were performed using the base and Argyle R-packages, and PLINK in a Unix environment (160, 161). We began with marker-level quality controls, because this was the order that was anticipated to conserve the most samples, while maintaining quality. The GigaMUGA array is thought to be slightly noisier than previous versions of this array, due to the high concentration of markers on the array, sometimes leading to signal saturation or bleeding of signal between samples. This leads to failing call quality or low call rates in a large number of the markers. It is important to remove these markers before assessing the quality of samples, so that samples are not penalized for containing large numbers missing calls for low-quality, unusable markers.

Individual genotype calls with Gen Call scores < 0.05 were marked as missing in Genome Studio by Neogen, before data was received (https://support.illumina.com/array/array_software/genomestudio/downloads.html). Gen Call scores are calculated by a Genome Studio proprietary algorithm, which incorporates distance of points on intensity value plot (X intensity by Y intensity) from the reference call cluster to determine the quality of each call. The 10% and 50% marker-level Gen Call scores are represent the Gen Call score value at the 10^{th} and 50^{th} percentiles of the score distribution of all calls ($N_{Calls}=N_{Samples}$) for each marker. Essentially, these scores represent the lower tail and median of the quality distribution for each marker, respectively. Our first step of quality control (QC) was to remove markers with: 10% Gen Call scores < 0.55 in both genotyping batches; or 50% Gen Call scores < 0.55 in either batch (Fig. 2). There were 27,391 markers that failed to meet the 10% Gen Call score threshold, 25,065 of which failed to meet the 50% Gen Call score threshold. An additional 25 markers failed to meet the 50% Gen Call score

threshold, leading to a total of 27,416 (27,391+25) markers being removed, with 115,843 markers remaining.

The second step of our QC pipeline was to remove markers with a low minor allele frequency (MAF), meaning the frequency of the less common allele in your sample. Rare alleles are hard to differentiate from technical error, and there is typically not enough power to detect their effects on the phenotype (162). Markers were removed if they had: a MAF of 0 in either genotyping batch; or MAF<0.02 when batches were combined. Because some of the "minor alleles" designated by Genome Studio were not necessarily the minor alleles (or alleles of lesser frequency for that locus) in our dataset, we also used an MAF=1 and MAF>0.98 (so that the frequency of the actual minor allele was 0 or 0.02, respectively) for these respective thresholds (Fig. 3). There were 7,745 markers with MAF = 0 or 1 in the first genotyping batch, 7,074 of which also had MAF = 0 or 1 in the second batch, and an additional 30 markers with MAF = 0 or 1 in the second batch. There were 295 markers with MAF < 0.2 and 3 markers with MAF > 0.98 when batches were combined, once MAF = 0 or 1 had been removed. This led to removal of a total of 8,142 ((7,745+30)+(295+3)) markers, leaving 107,701 markers remaining.

Finally, markers with a < 94% call rate in either genotyping batch were removed, based on spline of the left-hand tail of call rate histograms (Fig. 4). Low call rate indicates poor marker quality, possibly due to ineffective of oligonucleotide binding to DNA fragments. There were 1,661 markers in the first batch and 1,378 markers in the second batch with < 94% call rates, 765 of which overlapped between batches. This led to the removal of a total of 2,274 (1,661+1,378-765) markers, with 105,427 markers remaining (Fig. 4). Only 6 of these markers were on chrY, and 3,631 were on chrX. Because all of our samples are

males, heterozygosity in these markers, in non-pseudoautosomal regions, indicates calling error. Therefore, we removed sex chromosome markers for which > 10% (i.e. in > 63 samples) of the samples had heterozygous calls. This threshold was chosen based on the distribution of heterozygous call numbers (Fig. 5).

The remaining marker-level quality control step is to remove markers for which the allele distribution has an extreme deviation from Hardy-Weinberg equilibrium. However, the deviation statistic can be affected by relatedness between samples. Explained in more detail below, the process of determining relatedness required good quality samples. Therefore, this step was not performed until after sample quality control was performed.

Sample-Level Quality Control

The first step in sample quality control was to remove samples with low call rates, indicating overall low call quality. Samples with call rates < 98% were removed. A total of 21 mice were removed, 1 from the first microarray batch and 20 from the second, with 604 samples remaining. This threshold was chosen by visually determining the location of outliers in the overall distribution of call rates across samples (Fig. 6). Although this may seem like a large number of samples to remove, the inclusion of poor-quality samples actually decreases power due to noise. We therefore aimed for a stringent threshold that demarcated outlying nature from clusters of high-quality samples (Fig. 6). Mice that fell below this threshold likely possessed other confounds of technical variation from other mice, due to their outlying nature.

We next tested for sample cross-contamination by examining the heterozygosity rates within samples. Samples with abnormally high numbers of heterozygous calls across the genome have likely been contaminated by DNA from other samples, creating heterozygous

calls where the original sample is actually homozygous. Because the genotypic distributions at each locus likely vary across DO populations coming from varying combinations of breeders and generations, it is important to determine a unique threshold for each study. For our population, a heterozygosity frequency > 0.41 was > 3 standard deviations above the mean, and was considered to be contaminated (Fig. 7). Only one mouse was removed at this step, leaving 603 samples.

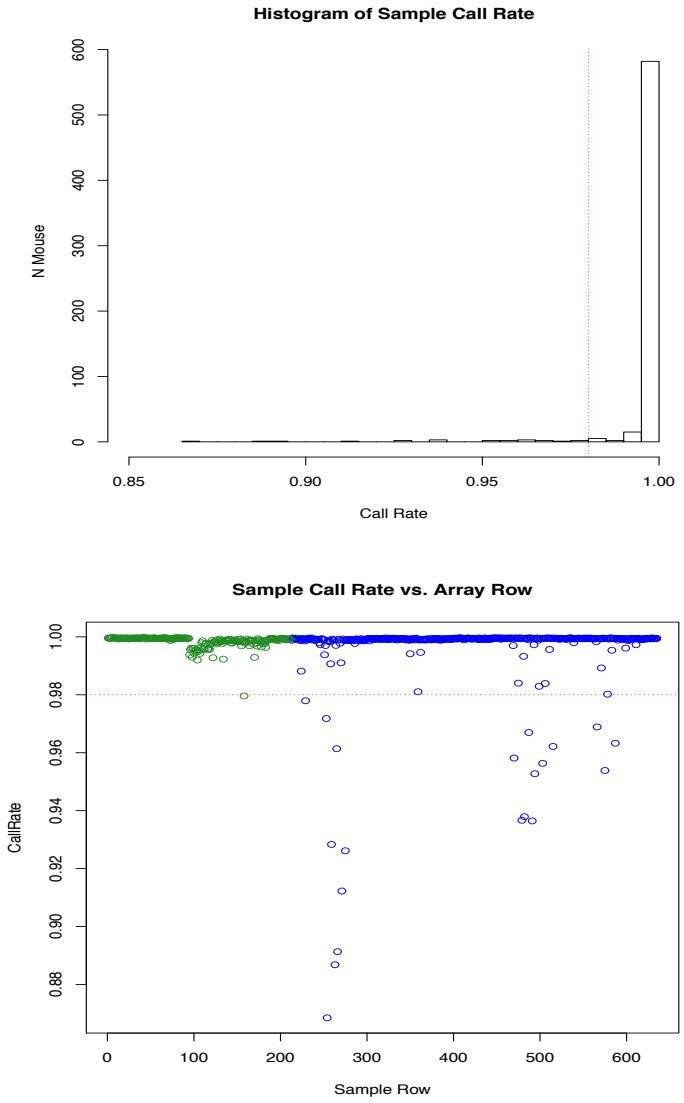


Figure 6: Histogram of sample call rates across all samples (top), and the relationship between call rate and the order in which samples were genotyped (bottom). Samples marked in green were genotyped in the first batch, and those in blue were genotyped in the second batch. The red dotted line depicts the same quality threshold seen in the left plot. Overall, the two batches had similar call rates, with minor deviations, although a slightly higher proportion of samples were removed from Batch 2 due to outlying low call rates.

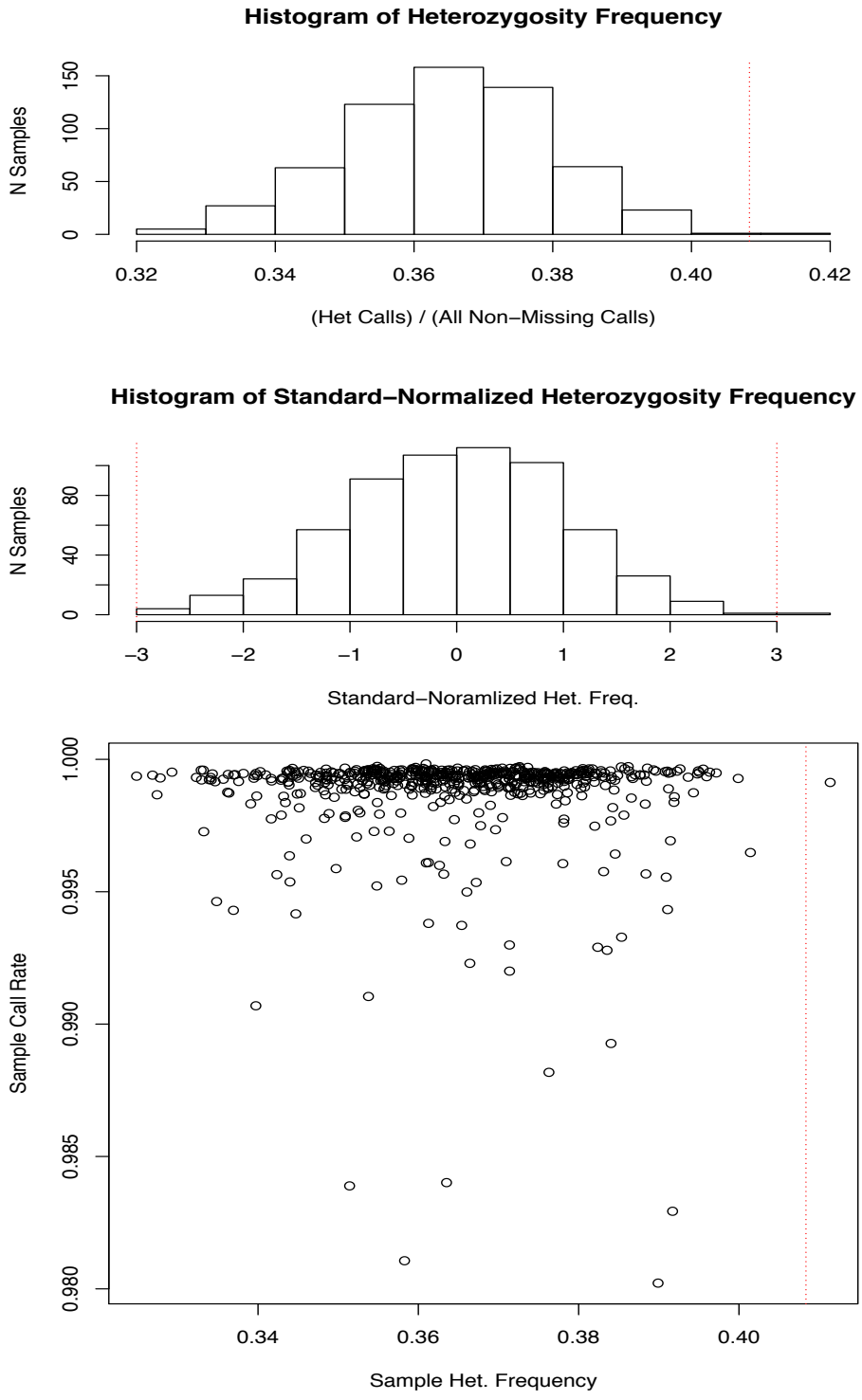


Figure 7: Histograms of heterozygosity frequency (or the proportion of each sample's total calls that were heterozygous) (top) and Z-normalized heterozygosity frequency (middle), and a scatter plot representing the relationship between heterozygosity frequency and call rate (bottom). Dotted lines in Plot B represent thresholds of + or – 3 standard deviations from the mean frequency. Samples falling outside of these quality threshold lines were removed. Dotted line in Plots A & C represents the raw heterozygosity frequency value that corresponds with the upper quality threshold in Plot B. No samples fell below the lower threshold. One sample fell above the upper threshold, indicating cross-sample contamination, and was removed. Call Rate did not seem to decrease as heterozygosity values became more extreme. Therefore, call rate would not have been a sufficient quality metric for sample-level quality control.

The final step of sample QC was to assure that all mice were genetically male, to assure there are no sex effects within our sample. Because of the relatively poor quality of the sex chromosomes, compared to the autosomes, we confirmed genetic sex using several different methods provided by the Argyle R-package (160). The first method was to identify true females based on the presence of high chrX marker heterozygosity paired with fewer than average high-quality chrY marker calls, or high chrX marker signal intensity and outlying low chrY marker intensity. No samples matched this description, and therefore all mice were considered to be genetically male. The second, more precise method was based on an algorithm in Argyle that calculates the probability of being male based on X and Y chromosome marker intensity. No mice appeared to have a much lower probability of being male than the other samples, and therefore all samples were still considered to be genetically male. For the final method, chrY calls were used to calculate the probability that these mice actually do not have a Y chromosome, but instead have two X chromosomes. Because of the lack of markers on the Y-chromosome after QC, all mice were called as "unknown", and this method was considered to be uninformative. All mice appeared to have been correctly sexed as male, so no samples were removed at this step.

As seen in Figure 8, overall sample quality was greatly improved by marker-level and sample-level QC. Overall missingness was lower, genotype distributions were more stable across samples, and intensity varied less across samples. (Notice the difference in the Y-axis scale between before and after plots). The genotype distributions were originally skewed, particularly in the samples with lower median signal intensity. However, after QC, the distributions appear as would be expected, with consistently low minor allele homozygous rates and proportionately higher heterozygous and major allele homozygous rates.

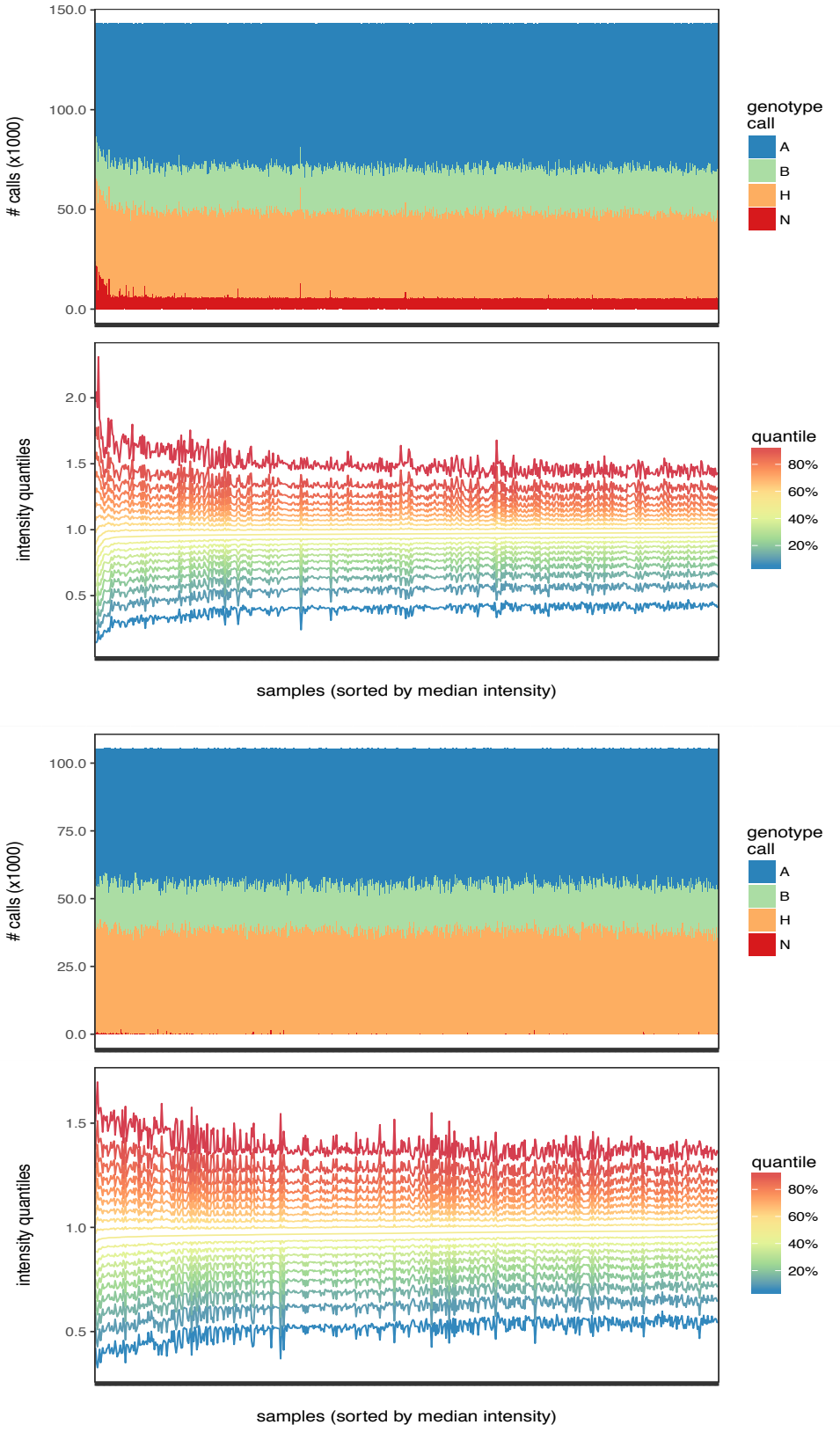


Figure 8: Stacked genotype call numbers ($N_{markers}$/sample) and linearly plotted raw intensity value quantiles for samples before (top) and after (bottom) marker- and sample-level quality control. Red shading and lines represent missing calls, orange represents heterozygous calls, and blue and green represents major and minor allele homozygous calls, respectively. Overall, missingness was reduced, and extreme intensity values were removed (notice the change in Y-axis range).

Sibling Identification and Hardy-Weinberg Equilibrium Test

Severe deviation from Hardy-Weinberg Equilibrium, or the expected distribution of genotypes in a randomly mating population at loci unaffected by natural selection or genetic drift, indicates genotyping error. In order to follow the expectations of Hardy-Weinberg Equilibrium, genotypes must have a distribution close to $p^2+q^2+2pq=1$, where p = minor allele frequency, and q = major allele frequency. However, as referenced above, high relatedness between samples can lead false positive deviations. Therefore, it was important to remove siblings from our population before identifying markers that do not satisfy this distribution. Although pedigrees for these mice were not available, it was possible to identify sibships via a relatedness metric, such as pi-hat estimated by PLINK (163). The *pi-hat* parameter indicates the estimated proportion of loci shared across the genome, with identity by descent, meaning they share the same alleles inherited from the same ancestor.

Based on distribution of *pi-hat* values across pairs of mice within each breeding generation of our sample, and with Identity By State (IBS; or sharing of the same alleles inherited from different ancestors) information, we were able to distinguish what range of pi-hat values likely indicated sibship (Fig. 9). *Pi-hat* values were calculated in PLINK1.7, using hard genotype calls from array markers (164). The small peak to the far right of the distribution of *pi-hat* values for each generation is likely composed of sibling pairs. This mode of the distribution appears to start around a *pi-hat* value of 0.3 - 0.4. We tested *pi-hat* sibship thresholds between 0.3 and 0.38, by intervals of 0.01 by examining the structure of sibship structure. Where a cluster of siblings in the family network plot were completely interconnected (i.e. every node shares an edge with every other node in that network), it was assumed that estimated sibships were true positives (i.e. the sibships identified were likely

true sibships). Where sibling clusters were mostly interconnected but were missing some links, it was assumed that the missing links were likely false negative sibships (i.e. sibships that truly exist were not identified). These assumptions followed the logic that if mouse A is a sibling of mouse B, and mouse B is a sibling of mouse C, then mouse A should also be a sibling of mouse C. Where single links of sibship existed between clusters of siblings, or links of sibships existed between mice from different generations, it was assumed that those were false positives (i.e. sibships were identified where one did not exist). Where a mouse displayed sibships with most of the mice in a cluster, but not with all of the mice (ex. if a mouse displayed sibships with 4 of 5 mice in a cluster that all shared sibships with each another), those missing links were assumed to be false negative sibships (Fig. 10). The ideal pi-hat threshold for determining sibship would balance the number of false negatives and false positives, indicating that most identified sibships were true positives, where lower thresholds would lead to a larger number of false positives and a higher threshold would lead to a larger number of false negatives.

A *pi-hat* value of 0.34 seemed to best meet these ideal conditions, and also did well at identifying the clusters of likely-siblings when comparing IBS values (Fig. 10 & 11). Therefore, pairs of mice with a *pi-hat* > 0.34 were considered to be siblings. Because 518 mice were identified as having siblings, it was necessary to keep one sample per sibling cluster, in order to remove highly related pairs while maintaining a large enough sample size to analyze Hardy-Weinberg equilibrium adherence. The mouse with the highest average within-cluster *pi-hat* estimate for each sibling cluster was kept, as they were thought to be the best genetic representation of the whole cluster. There were 32 sibship clusters and 8 unclustered mice in generation 22, 39 clusters and 23 unclustered mice in generation 23, 34

clusters and 18 unclustered mice in generation 24, and 53 clusters and 33 unclustered individuals in generation 25. This led to a total of 240 mice for estimation of Hardy-Weinberg equilibrium deviation. Because the remaining mice were still more interrelated than a population of unrelated humans, a very lenient threshold of $p < 5x10^{-10}$ was used to specify markers with extreme deviation from Hardy-Weinberg equilibrium. This led to the removal of 3,679 markers, leaving us with 101,717 markers.

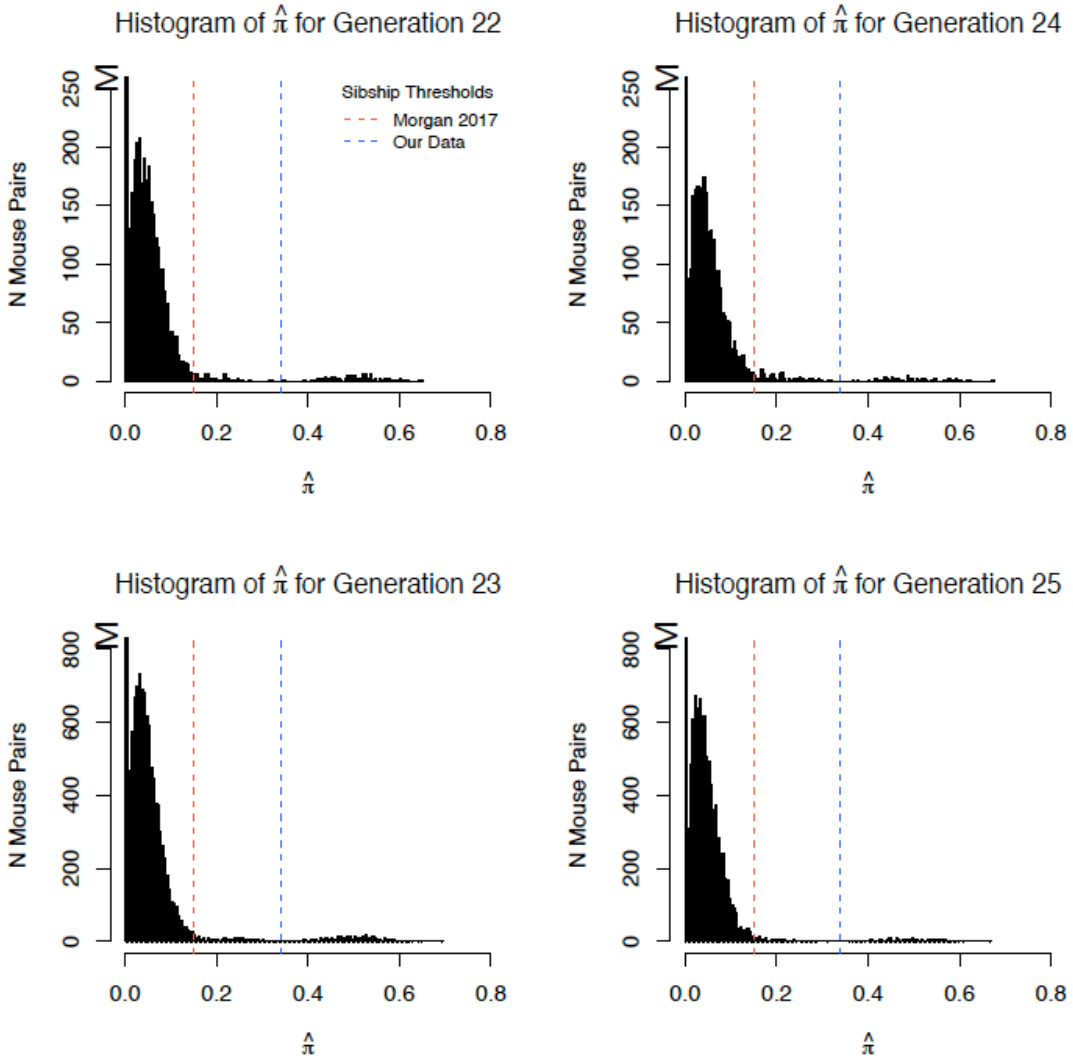


Figure 9: Pi-hat histograms for each possible pair of mice within each generation of mice, with our chosen threshold (blue) and the threshold published in Morgan et al., 2017, above which pairs are defined as siblings. The large disparity between our thresholds shows the importance of redefining such parameters according to each study's sample.

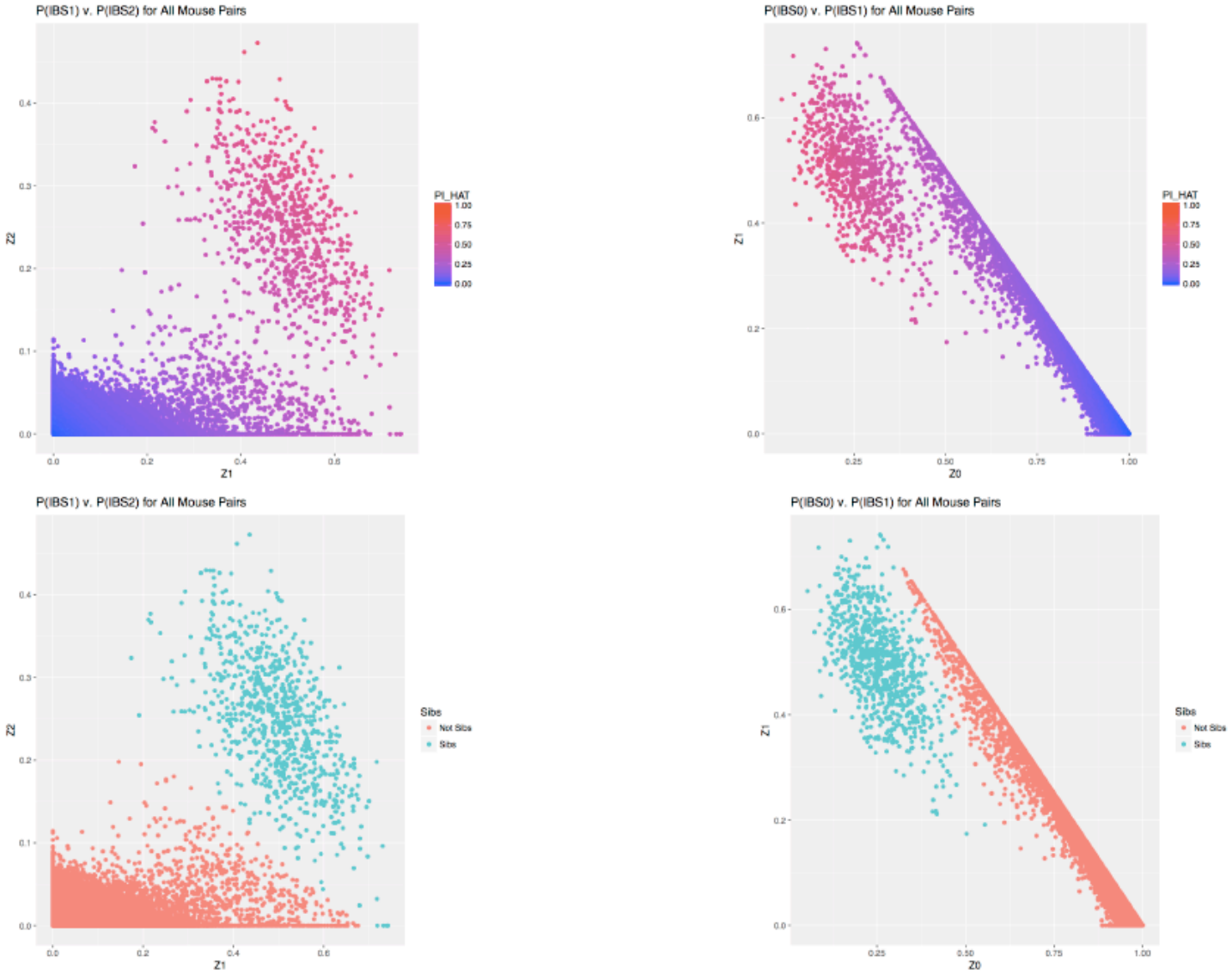
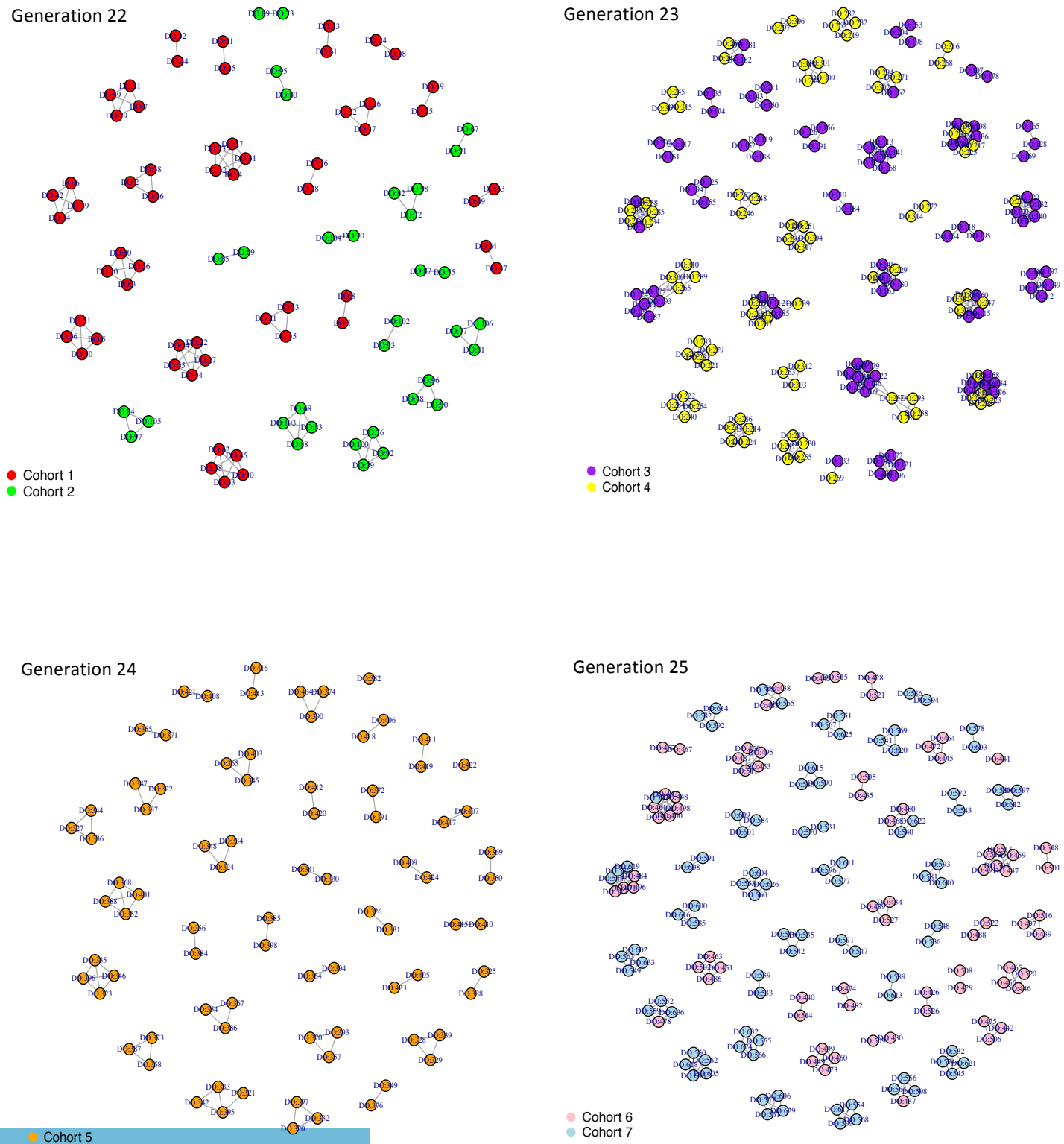


Figure 10: Probability of sharing 0 (Z0), 1(Z1), or 2 (Z2) alleles Identity by Decent (IBD) for any locus in the genome, for each possibly mouse pair. Color represents pi-hat values (top), by which sibships were defined, or by sibship status (bottom), defined as pairs with pi-hat > 0.34. As should be seen, siblings cluster closer to the upper right quadrant of the left-most plots and towards the upper left quadrant in the right-most plots, indicating accurate identification of sibship with pi-hat threshold 0.34.

a



b

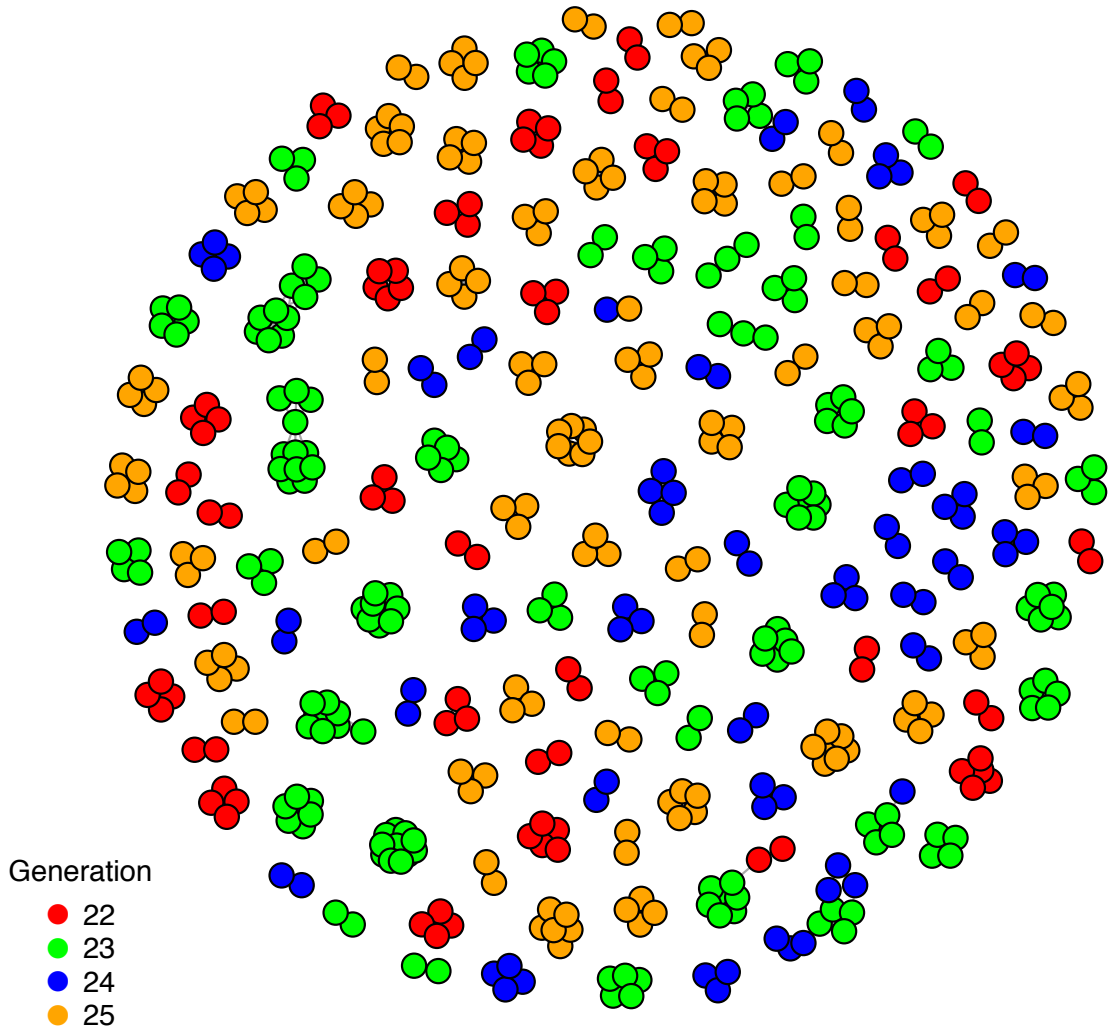


Figure 11: Sibling clusters, for which sibships are defined as mouse pairs with pi-hat > 0.34, for each generation of mice (a) and across all generations (b). Each circle or node represents a single mouse, colored by experimental cohort (a) or generation (b), and each connection or edge represents a sibship. Incomplete interconnectedness within clusters (ex. Cohort 3 triad on far right of Generation 23's plot) indicates false negatives in sibship identification, and sparse connections between clusters (ex. connections between one cluster from each cohort seen on far left of Generation 23's plot) or between generations (ex. bottom right blue and green cluster of Plot B) indicate false positives.

Final Numbers and Generation Effects

Before analyses began, seven additional mice were removed due to potential sample mix-up during data collection documented in laboratory notes. This left a final number of 101,717 markers and a total of 603 mice, 554 from the treatment group and 49 controls. Because markers on the Y chromosome were few and likely unreliable, it was not assessed in the analysis. The X chromosome was included in the analysis but was separated from the autosomes for permutations, and results were interpreted with caution, as fewer than 20 markers remained after QC to be used for imputation. Both sex chromosomes were excluded from kinship calculations. This treatment of sex chromosomes is provided as an option in the R/qtl2 package's permutation function (165).

Choosing Covariates

Several potential covariates with genotype existed in this dataset, including: breeding *generation*, *litter order*, experimental *cohort*, experimental *room* (2 rooms per cohort), *age* (8 weeks vs. 12 weeks), and *genotyping batch*. We first wanted to ensure that there were no major effects of generation on genotypic similarity, aside from what might otherwise be accounted for by a kinship coefficient. Although the inclusion of a kinship coefficient corrects for familial relatedness between mice within generations, there may still be genetic differences between generations that are shared between unrelated individuals within generations. These differences could be due to differing levels of LD (decreasing with each generation of breeding), genetic drift, or even fertility-based natural selection (73).

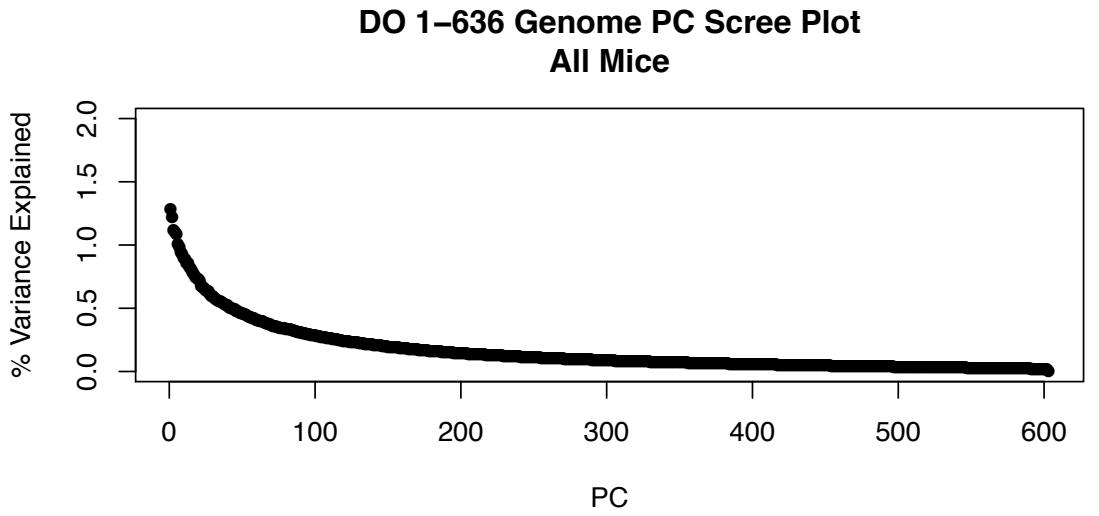
To visualize the effects that generation might be having on genotype, we performed principal component analysis for genotype calls across the entire genome, using PLINK,

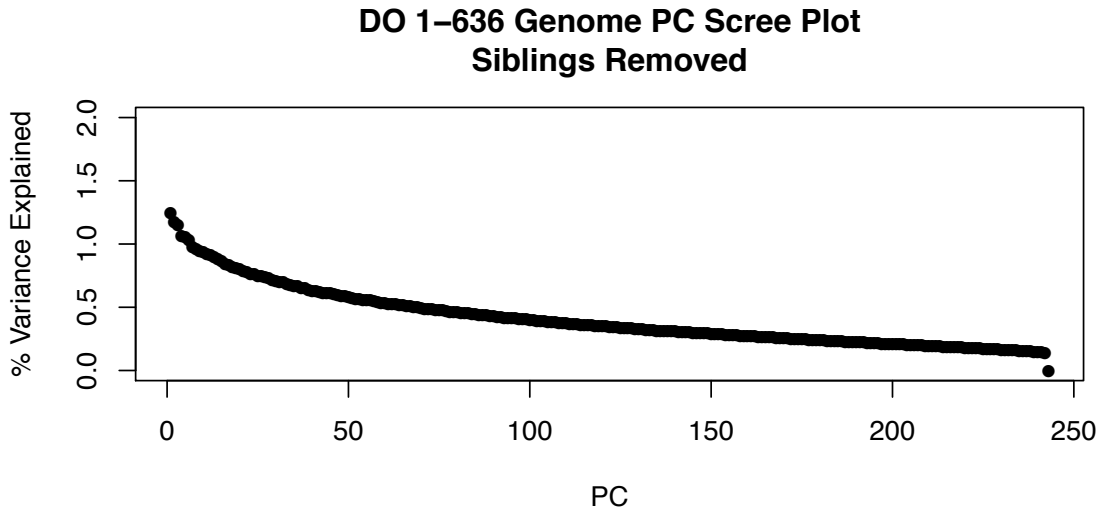
using the maximum number of principal components (PCs) (N_{PCmax}=N_{Mouse}=603). We performed the same analysis with siblings removed, in order to remove effects of true relatedness on generational clustering (N_{PCmax}=N_{Mouse}=243). (The number of sibships was slightly lower after removing Hardy-Weinberg-failing markers, leading to 3 additional samples after removing sibships.) Scree plots indicated that the first 3 PCs explained slightly more than 1% of the overall variance, each (Fig. 12). The projection of values onto each of the first 3 PCs was plotted against each other, and points were colored by generation, to visualize generational clustering (Fig. 12). Before removing sibships, it appears that there are 3 clusters (2 along both the 1st and 2nd PCs, and one along the 3rd PC) of mice from generation 23 (and some from generation 24) that are genetic outliers. These groups seem to dissolve into a single point for each cluster, once sibships were removed. This indicates that there were actually 3 outlying families in generation 23 (and 3 cryptic relatives from generation 24), as opposed to generation 23 being genetically distinct from the other generations, overall. In fact, each of these groups represented clusters of closely related sibling groups, with a likely cryptically related sibling cluster from generation 24, that are less closely related to the overall sample than most other groups of siblings (Fig. 13). Therefore, we concluded that the outlying nature of these mice is likely due to family structure, as opposed to generational differences, and inclusion of a kinship covariate in analyses would correct for these differences, with no further correction needed for generational differences.

All of the covariates mentioned at the beginning of this section are highly correlated with experimental cohort, and could therefore not all be included in the analysis together. For this reason, we aimed to identify the covariate(s) that accounted for the largest amount of

variation in ethanol consumption. To do this, we performed linear regression and/or ANOVA (depending on the nature of the covariate) to test for linear and nominal (or group-wise) effects of each covariate, and their interactions, on first-week, last-week, and whole-study mean total ethanol consumption. Ordinal covariates, such as cohort and generation, were tested as both continuous and categorical variables, to examine linear and group-wise effects. When tested by themselves, the following covariates had significant ($p<0.05$) effects on ethanol consumption for at least one time range: *generation*, *weight*, *genotyping batch*, and *cohort*. *Generation*, *genotyping batch*, and *cohort* were collinear with each other, and therefore could not be included as independent variables in the same analysis. Because *cohort* has 7 levels (compared to the 2 levels for genotyping *batch* and 4 for *generation*) over which consumption significantly varied for all time intervals tested, it is the most nuanced of the three covariates, while encompassing the effects of the other two, as well. *Cohort* needs to remain in the model, because it captures experimental, microarray batch, and generational variance. Therefore, we chose *cohort* as the covariate to carry forward for QTL analysis. Although there was an overall significant negative linear relationship between *cohort* and ethanol consumption, it is not a directly linear relationship (meaning not all cohorts drank more than the *cohort* that directly followed them). In fact, cohorts 6 and 7 drank more than cohorts 4 and 5 at all time intervals. Therefore, cohort was operationalized as a categorical variable in QTL analysis by creating 6 contrasting binary variables (representing membership (1) or non-membership (0) for each of cohorts 2-7), using the first cohort as the reference (all 0s), meaning each cohort's effects were measured in reference to its difference from the first cohort.

a





b

Figure 12: Scree (a) and principal component (b) plots from principal components on all samples (top) vs. after siblings were removed (bottom). Colors in (b) represent breeding generation, and axes represent projections of genome-wide genotypes for each mouse on each of the first 3 principal components (PC1, PC2, and PC3).

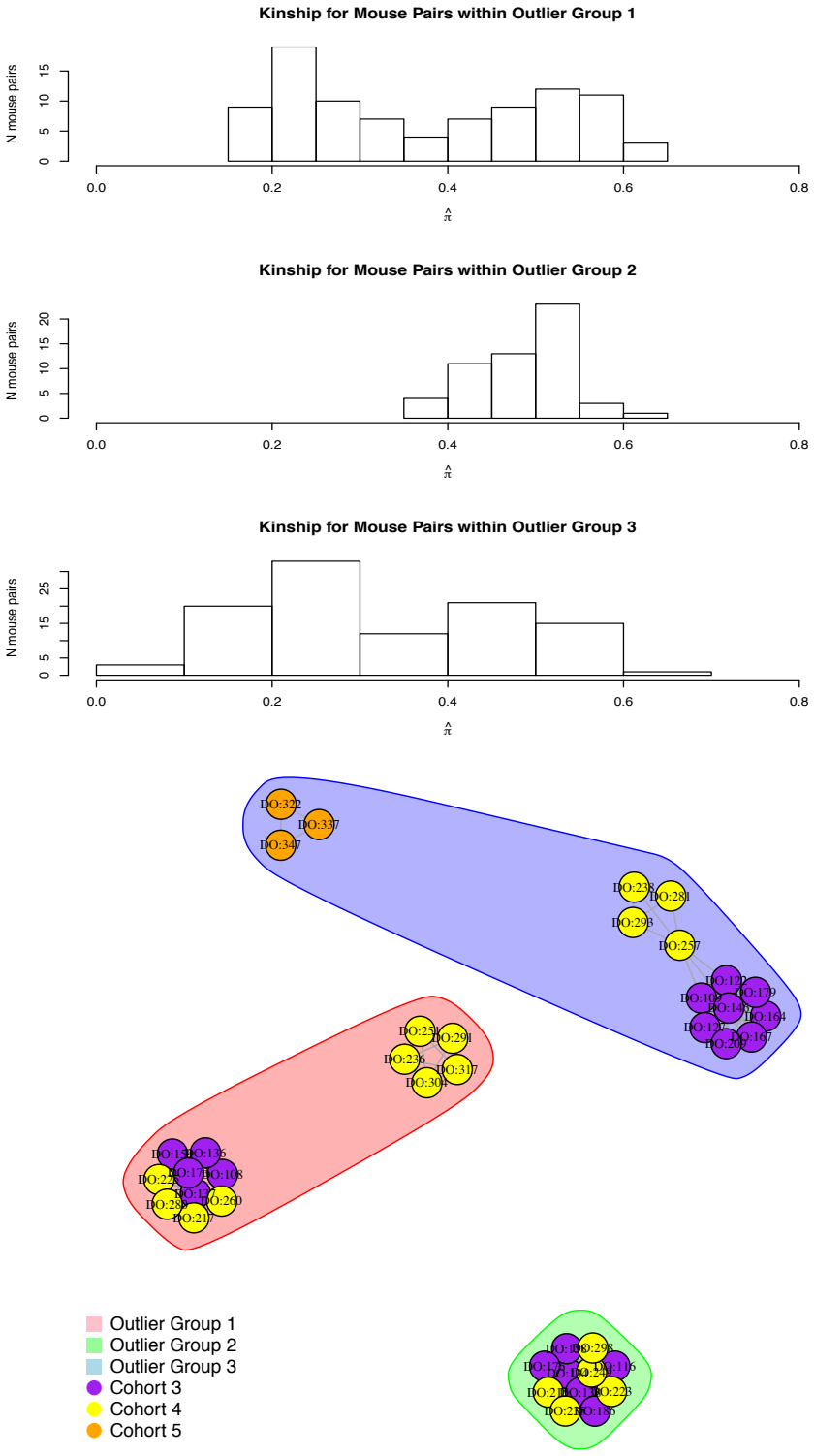


Figure 13: Histograms of pi-hat values for each possible pair of mice within each outlier cluster from principal component analysis on genome-wide genotypes (top). Bimodal distributions, as seen in Groups 1 and 3, indicate the presence of multiple levels of relatedness within a group. Unimodal distributions, seen in Group 2, with a lower pi-hat limit greater than the sibship threshold (0.34), indicate the presence of sibships only. Network plots of clusters (b) show clusters of siblings and cryptic relatives or first cousins (multiple clusters, either sparsely connected or unconnected with each other, within each group) for each group of outliers. Node colors represent experimental cohort, whereas background color represents outlier group membership.

Behavioral QTL Mapping

Genotype calls were converted to progenitor haplotype probabilities (one probability per progenitor strain at each locus) for QTL mapping. SNP probabilities, used for SNP mapping within QTL support intervals, were then imputed at intervals of 0.01Mb, based on these haplotype probabilities.

For each bQTL analysis, we applied linear mixed modeling via the R/qtl2 package (165), which contains functions catered specifically to QTL mapping in DO populations. The kinship matrix, representing the pairwise relatedness between all mice, was calculated using the calc.kinship() function in R/qtl2, using the Leave One Chromosome Out, haplotype-based method. Kinship was included as a random predictive variable in all of our analyses, to preclude false positives attributable to increased behavioral similarities between more highly related mice. Because generation covaried linearly with cohort, but cohort captured more variance, cohort was included as the only covariate (aside from kinship) in our analyses. Phenotypes were either log- or square-root-transformed to obtain normality, before running analyses. The QTL analysis for each trait was run on haplotype probabilities, resulting in 7 degrees of freedom. Because of the computational intensity of SNP analysis using R/qtl2 given the large number of SNPs within the DO genome, SNP associations (with 1 degree of freedom) were only calculated for all imputed SNPs within the 95% Bayesian support interval of each significant or suggested locus.

Selecting Kinship Estimation Method

Before selecting a method of estimating kinship, we compared bQTL results between analyses that incorporated relatedness matrices of the following types: haplotype probability-based kinship; haplotype probability-based kinship, leave one chromosome out (LOCO);

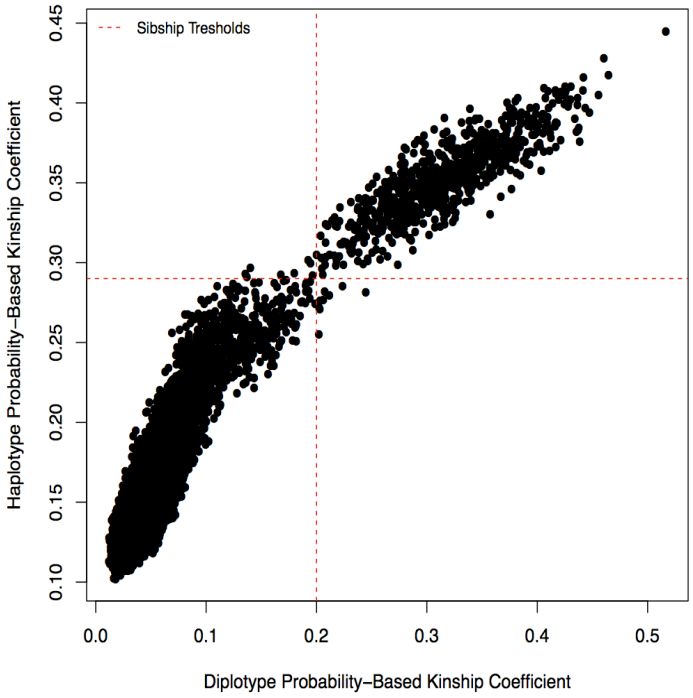
diplotype probability-based kinship; diplotype probability-based kinship, LOCO; Genomic Relatedness Matrix (GRM). The former 4 types of kinship were estimated using the qtl2 package, based on progenitor haplotype and diplotype probabilities at each locus, respectively. This was done using the first genotyping batch of mice (Experimental Cohorts 1 & 2, N=210) for *Coat Color*, expecting to see a peak at the known contributing locus on Chr7. The GRM was estimated using standard methods in PLINK1.7 (161). In order to determine the relative quality of each method, with respect for correcting for relatedness between samples to reduce the false positive rate, we compared the size of the LOD peaks at the known *Coat Color* loci and the noise for each set of results. The signals within the expected regions (Chr7: 87.43-87.49Mbp for albinism, and Chr2:154.79-155.05Mbp for black coat color) were considered to be true positives, whereas all other signals were considered to be noise. For these two separate analyses, albinism was considered to be a binary trait (due to it's recessive nature); and black coat was considered to be a categorical variable with 1 indicating albinism, 2 indicating an agouti coat, and 3 indicating a black coat (due to this locus's co-dominant nature). The number of noise-attributable loci was determined by the number of LOD peaks with a top score of at least LOD=4 and a LOD-drop of 1.5 on either side of the peak. The level of real signal was determined by the top LOD score of the peak that fell within the expected interval. The goal was to achieve the highest signal with respect to noise (i.e. the largest difference between mean true signal LOD scores and mean background LOD scores).

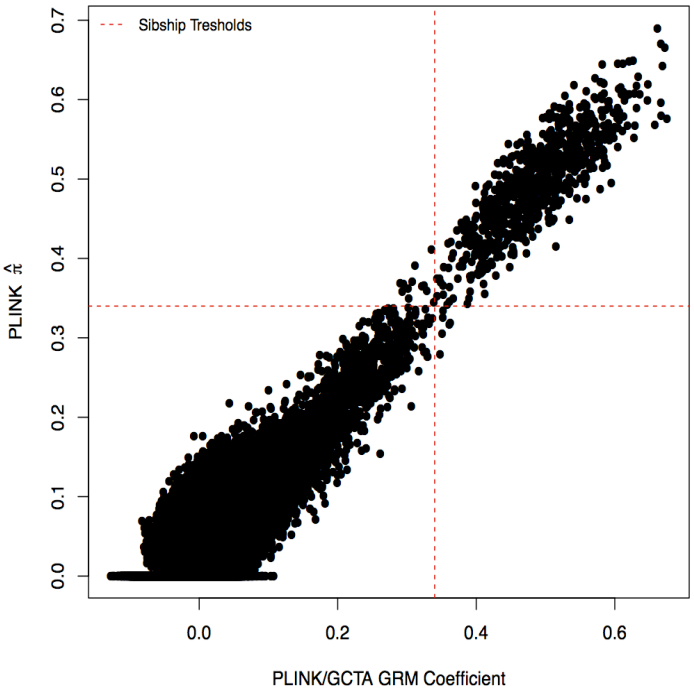
The inclusion of diplotype-based kinship estimates in the model led to the highest peak LOD scores, but also the highest noise, indicating underestimation of relatedness, which imparts the potential for a higher rate of false positives in behavioral analysis. Similar results

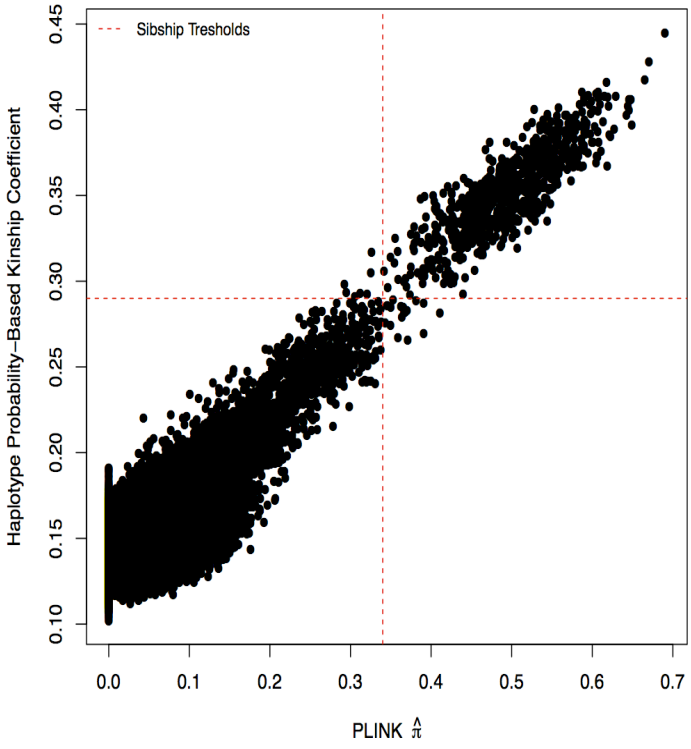
were seen when using GRM in place of kinship. Haplotype-based method of kinship estimation seemed to lead to the highest LOD peaks while maintaining relatively little noise, indicating that it balanced the false and true positive rates better than the other methods. To compare relatedness coefficients directly, we plotted out the histograms of values for each method and determined sibship thresholds based on the location of the beginning of the second curve of the bimodal distributions (as described for the first step of sibship determination for quality control purposes) (Fig. 14). We then plotted the values estimated by each method against each other, to identify relative over- or under-relatedness (Fig. 14). Keeping in mind that the X -and Y-axes are of different ranges, it appears that the diplotype-based kinship estimates are lower than the haplotype-based estimates for non-siblings, but higher for siblings. The same is true for the GRM values. However, the GRM values were lower than the diplotype-based estimates, particularly for non-siblings. Although the GRM values may not be directly comparable to the other two estimates (because the GRM was estimated using original genotype hard-calls, and the other two were based on estimated progenitor haplotype or diplotype probabilities), the overall distribution of comparisons is informative.

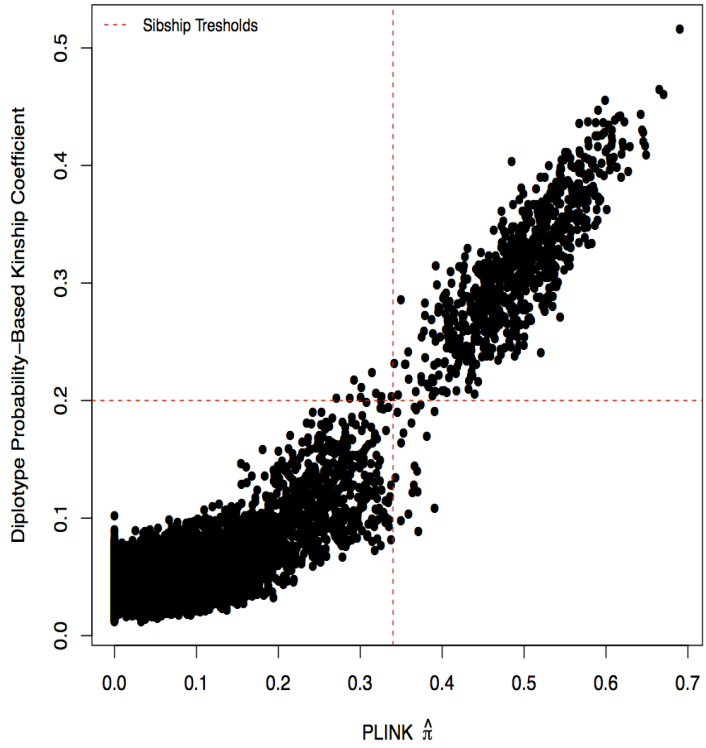
Based on the QTL results and the kinship estimate value comparisons, it appears that GRM values underestimate relatedness between "unrelated" (as unrelated as possible in this type of population) mice and overestimate relatedness between siblings (Fig. 14-15 & S1). This means that it is overcorrecting for the differences in similarity between related and unrelated mice, masking both false-positive and real signals. The same is true for the diplotype-based estimates, but to a lesser extent. These results indicate that the haplotype-

based estimation method is best suited for this type of analysis, balancing specificity and sensitivity better than the other two types of estimates.









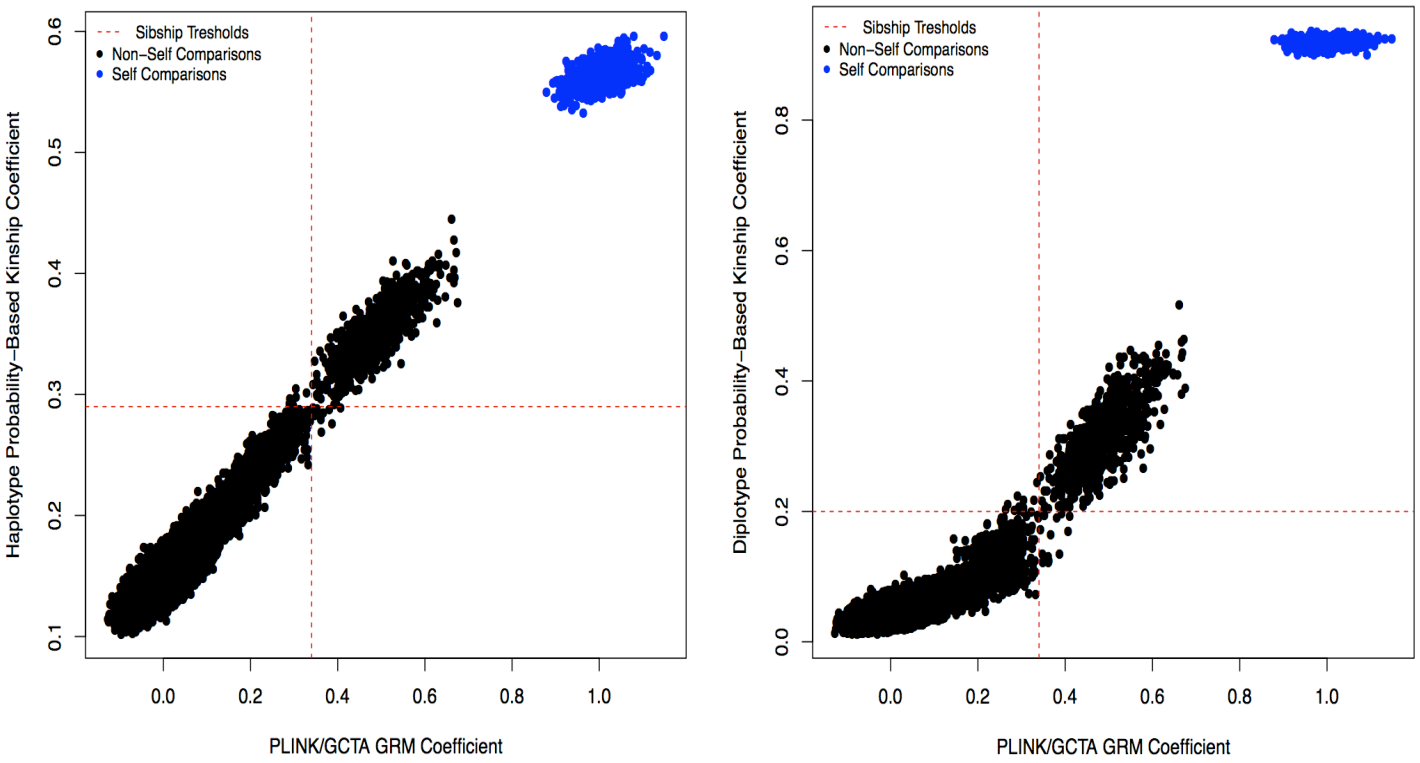
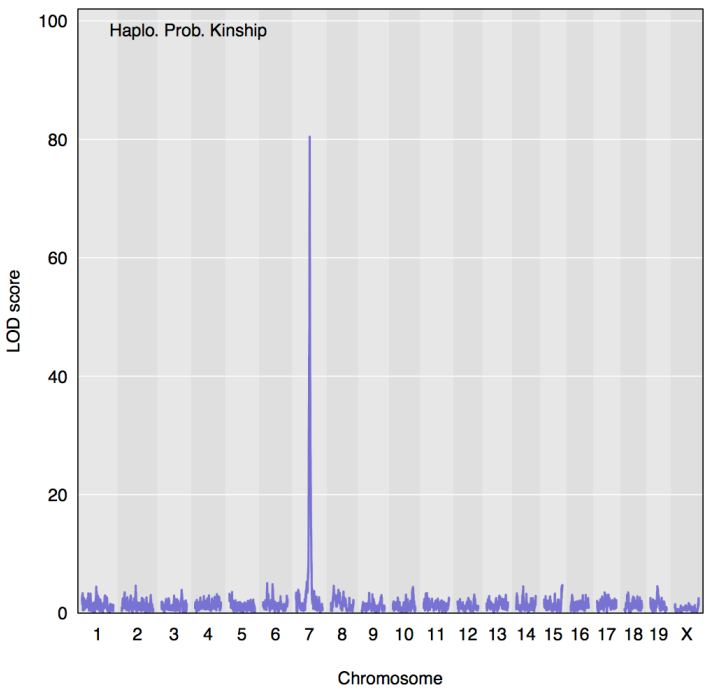
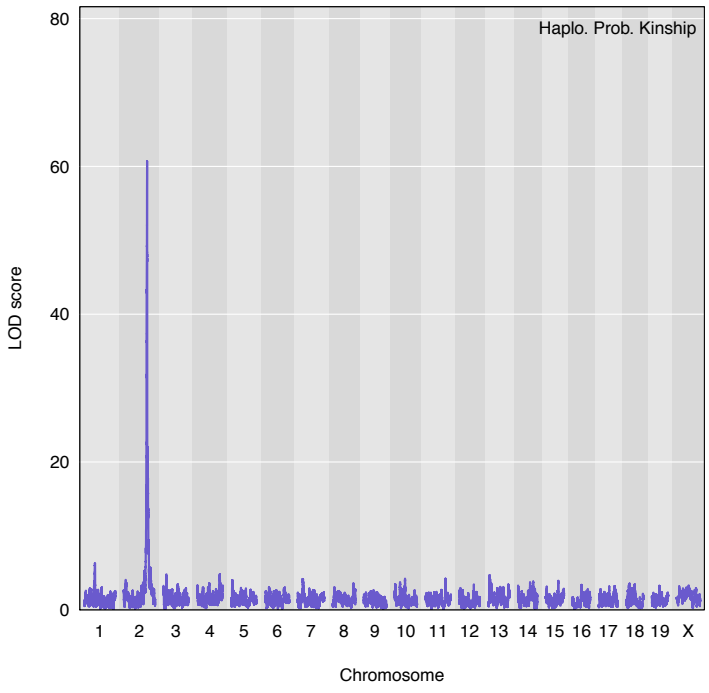
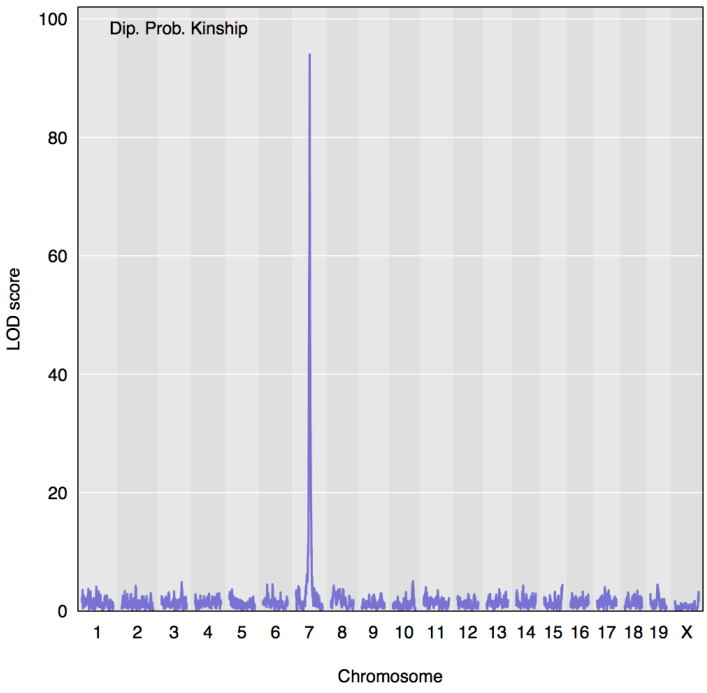


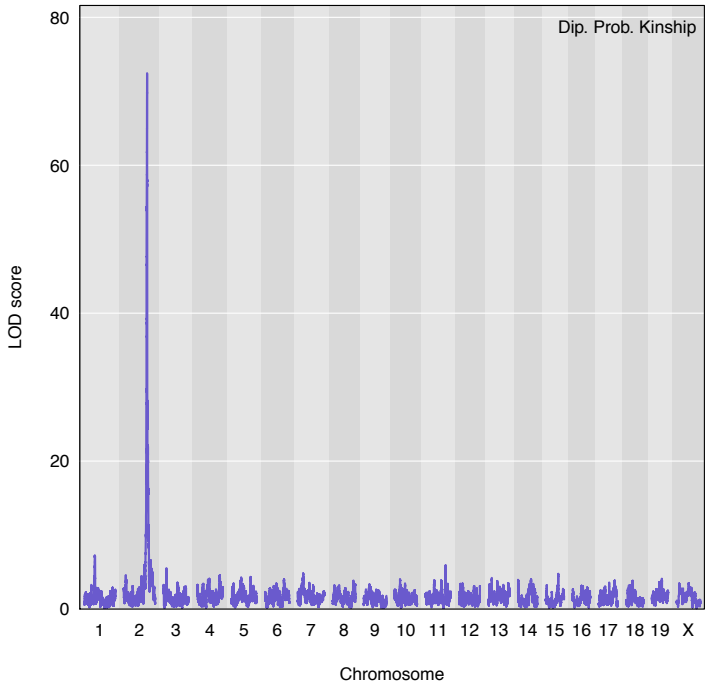
Figure 14: Scatter plots comparing relatedness metrics for each possible pair of mice, from each estimation method: pi-hat, Genomic Relatedness Matrix (GRM), haplotype-based kinship, and diplotype-based kinship. Self relatedness values are missing from plots containing pi-hat and GRM estimates, as those estimates are not calculated for these metrics in PLINK. Where self relatedness was calculated (can be < 1 due to weighting by probability of allele sharing estimates), these values are colored in blue. Dotted red lines indicate what would be used as sibship thresholds for each metric, above which pairs are likely siblings.

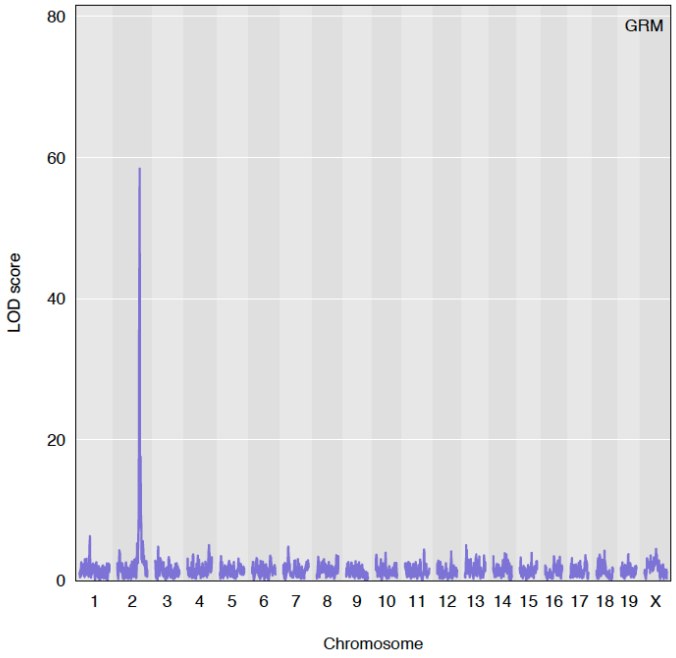
a

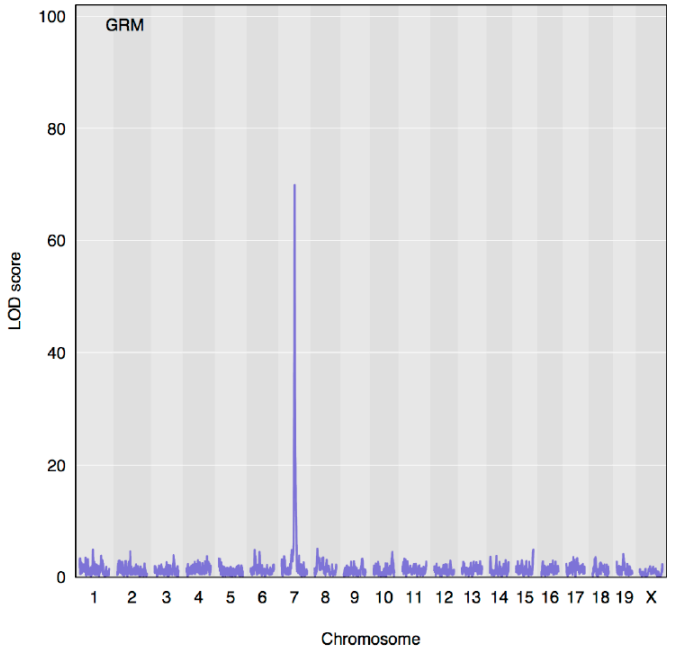


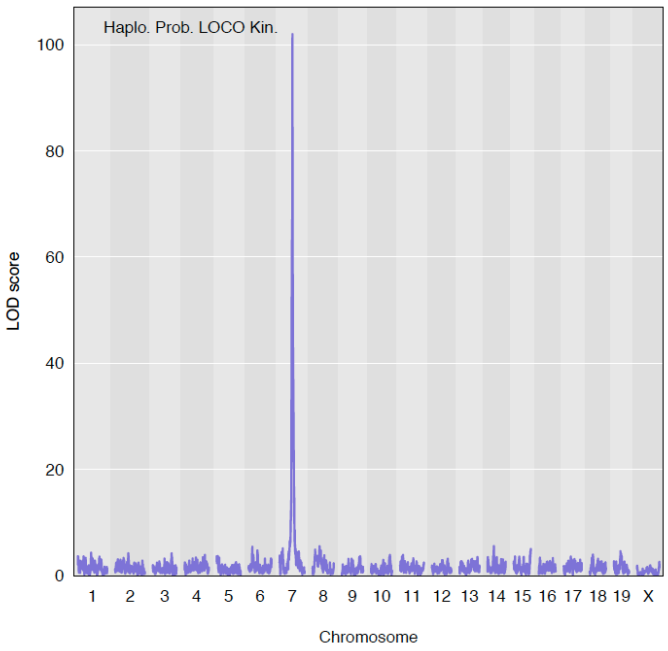


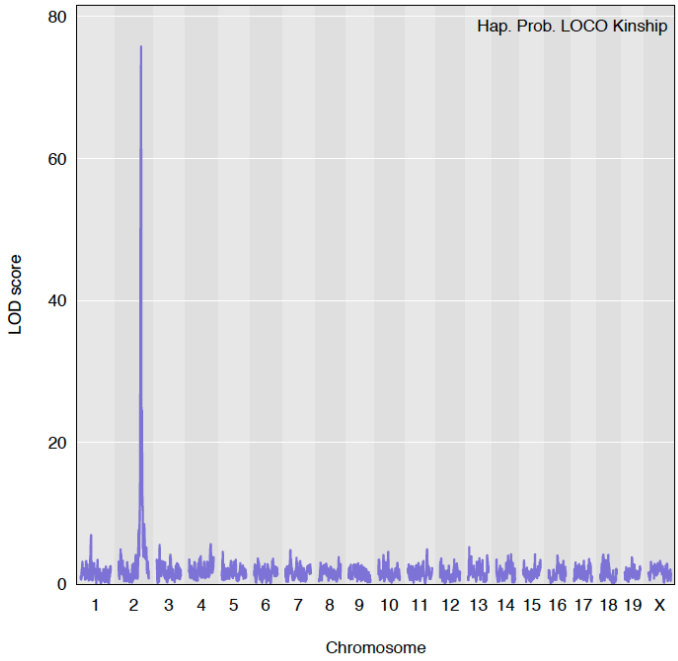


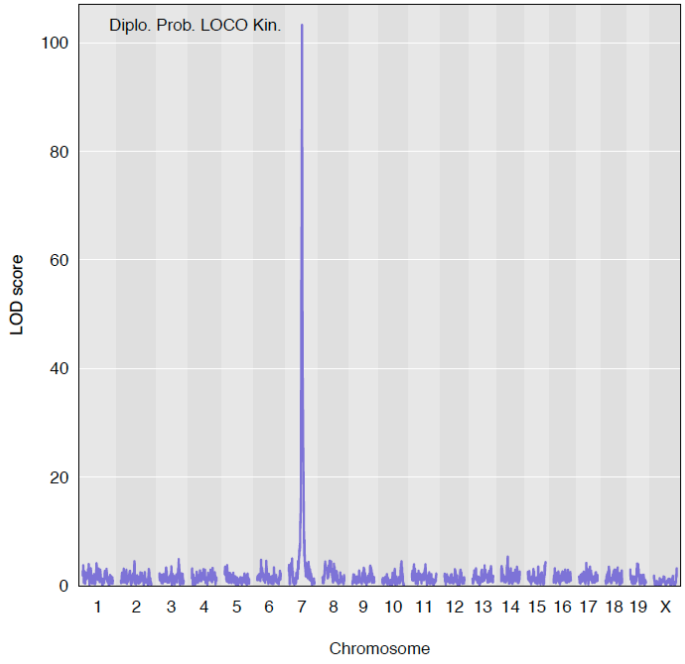


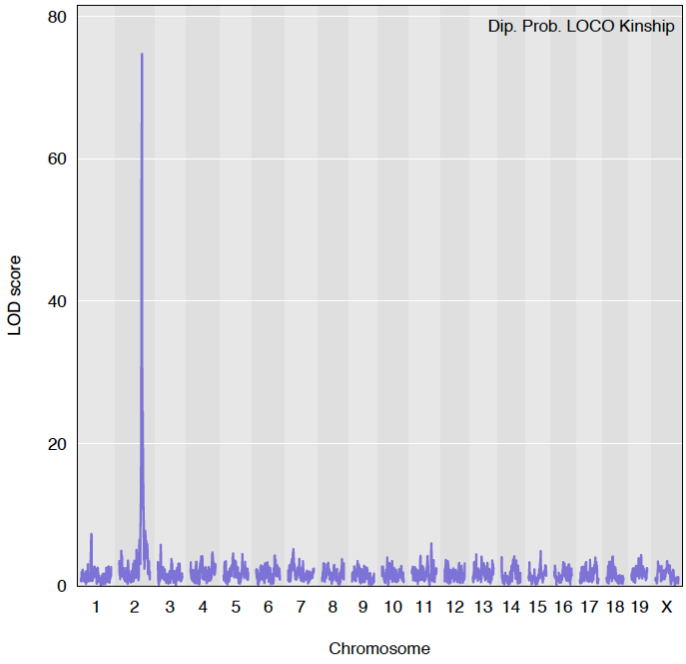












b

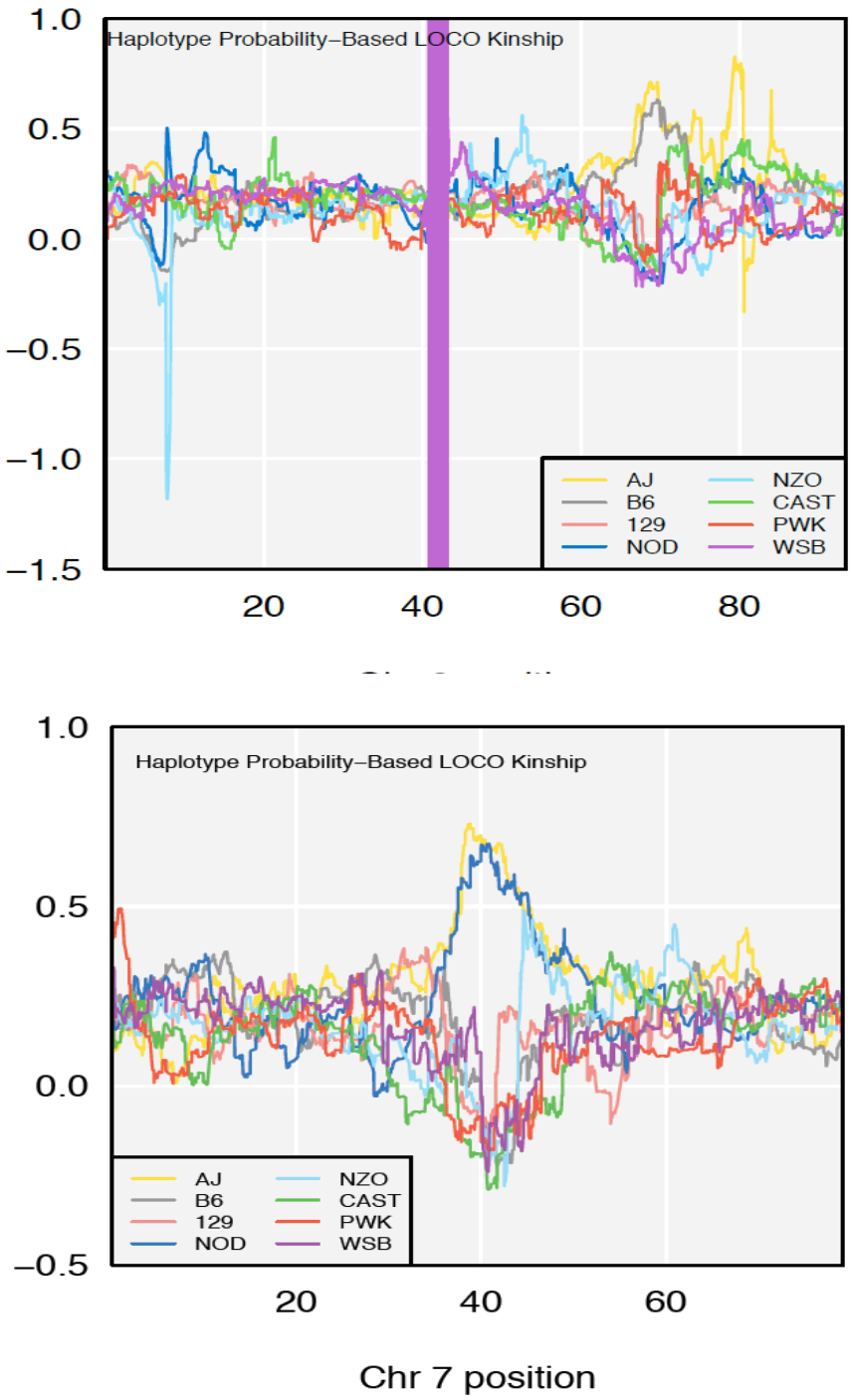


Figure 15: (a) LOD plots for the albino locus on Chr7 (left) and the black coat color locus on Chr2 (right), for separate analyses using each type of relatedness estimation (Diplotype Probability-Based Kinship, Haplotype Probability-Based Kinship, Genomic Relatedness Matrix (GRM), Diplotype Probability-Based Leave One Chromosome Out (LOCO) Kinship, and Haplotype Probability-Based LOCO Kinship) as a random variable covariate, across all chromosomes. Peaks on all other chromosomes are likely detecting noise. (b) Founder effect plots for these respective analyses. Each line represents the average phenotypic value (although traits were binary, only linear regression methods were available in the R/qtl2 package) for mice in which the haplotype in that region most likely came from the same founder, with contributing founder strain denoted by color. Vertical purple shading on Chr2 represent areas in which the frequency of the WSB haplotype has been fixed to 0, due to breeding complications, leading to zero genotypic variance in that region. R/qtl2-calculated kinship metrics incorporate this information into the model, whereas GRM does not.

As explained in Cheng et al. (166), the LOCO methods for estimating haplotype- and diplotype-based kinships increase signal. This method leaves the chromosome that is currently being analyzed out of the kinship estimation, so that it does not account for differences attributable to loci on that chromosome, which would lead to false negative results. In order to ensure that the LOCO method maintained a good signal-to-noise ratio, we compared the haplotype- and diplotype-based estimated with and without LOCO. The LOCO estimates led to much higher signals with only slightly higher levels of noise, compared to the whole-genome kinship estimates. We therefore chose to use the haplotype probability-based LOCO kinship estimates as the relatedness covariate in our bQTL analyses. The location of peaks relative to the know location of the effective loci are depicted in Figure 16. These results show that our data correctly estimated the locations of these QTL, with very few peaks more than 8Mbp away from the true signal. Furthermore, the founder effects were correctly estimated (A/J and NOD haplotypes conferring albinism and C57/BL6J and A/J haplotypes conferring black coat color) by our data (Figure 16). (A/J mice are homozygous at both loci, but the homozygous recessive albino phenotype is dominant over the black coat gene.) These results indicated accurate, clean genotypic data, and that there have been no large sample mix-ups.

Table 1: Permutation-Derived LOD Thresholds for QTL Significance

Phenotype	LOD Threshold (p<0.05)
Whole Study Tot. EtOH Cons.	7.470
First Week Tot. EtOH Cons.	7.546
Lasat Week Tot. EtOH Cons.	7.691
Whole Study Tot. EtOH Pref.	7.535
First Week Tot. EtOH Pref.	7.517
Lasat Week Tot. EtOH Pref.	7.573
Whole Study 30% EtOH Choice	7.700
First Week 30% EtOH Choice	7.470
Last Week 30% EtOH Choice	7.488

LOD scores corresponding to the 5th percentile of the permuted LOD score distributions, for each tested phenotype. Values greater than these thresholds were considered to be significant.

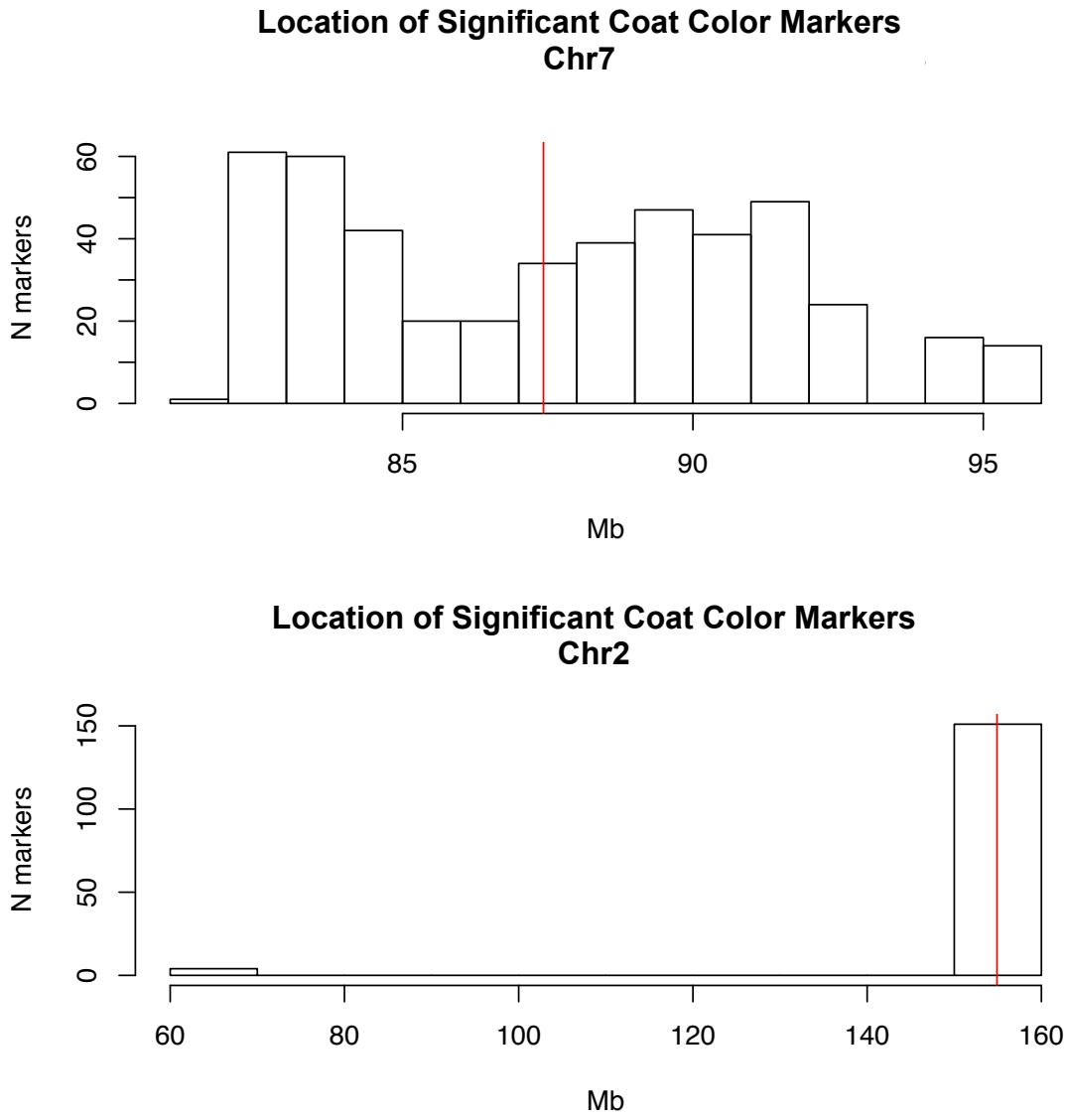


Figure 16: Slide 35: Histogram of signal peak locations with LOD > 5 for each QTL peak identified for albino (top) and black (bottom) coat color, on chromosomes 7 and 6 in units of megabases, respectively. Red lines indication the location of the actual SNPs known to affect these phenotypes. Overall, most peaks were within 7Mb of the real affective loci for albinism, and within 5Mb of that for black coat color. All peaks with LOD > 5 identified for either phenotype are shown here. These results indicate that genotypic data is accurate and of good quality, following quality control.

Significance-Level Calculations

Empirical p-values were calculated from LOD score distributions created from one thousand permutations on randomized samples to determine, for each phenotype, using the scan1perm() function from the qtl2 package. Phenotypes were randomized across samples, as opposed to genotypes, to maintain linkage and kinship structure. For each permutation, the highest LOD (logarithm of the odds) score across the genome is stored, creating a final distribution of 1,000 highest LOD scores from random samples. An empirical p-value < 0.05 was considered to be significant (Table 1). Because this method of permutation is highly conservative, a threshold of LOD=6 was chosen to identify suggestive peaks, roughly corresponding to the conventionally used suggestive threshold of $p < 0.63$ on our permuted distributions and representing peaks that could clearly be visually distinguished from noise in LOD plots (35, 167).

Founder Effect Calculation and Application

Within each suggestive or significant QTL region, founder effects, or the general contribution of each founder strain at loci across a chromosome to the QTL signal, were calculated using scan1coef() function in the qtl2 package. This function calculates the mean and standard error of a given phenotype for all mice that share the maximally probable contributing founder strain at each locus. Plots of these means, by contributing founder strain, allows the user to visualize which founder haplotypes confer higher or lower phenotypic values. This knowledge can contribute to the narrowing down of SNPs within each QTL region, by comparing the distribution of alleles at that SNP to the expected grouping for the effective SNP, based on founder effects. For example, if PWK and WSB haplotypes at a given locus confer higher ethanol consumption values than the other strains,

matching SNPs would be identified as those that possess alleles shared between, and unique to, PWK and WSB amongst all of the founder strains. No significance values are produced for the effect of each founder haplotype in a given region. Therefore, the expected founder allele grouping was determined by viewing the phenotype grouping in the founder effect plots (as described in the previous hypothetical). For ambiguous groupings, all possible groupings were considered. Consider another example in which PWK and WSB haplotypes conferred high ethanol consumption values, CAST haplotypes conferred moderate values, and the remaining strains conferred low values. In this situation, SNPs that satisfy the observed grouping patterns could have: a) an allele shared between PWK and WSB, and the other shared between CAST and all other strains; b) an allele shared between, and unique to, PWK, WSB, and CAST; or c) an allele shared between PWK and WSB, an allele unique to CAST, and a third allele (rare, but possible, especially when considering copy number variants) shared between the remaining strains. A function was written to check all SNPs for allelic distributions across the founder strains for matches to any one possible expected pattern, within each QTL support interval. Accepted patterns for each QTL can be seen in Table S1. Top-scoring SNPs were defined as those with a LOD score within 1.5 of the top LOD score, or those that were not analyzed (i.e. did not have enough information or were not imputed) but fell within the interval between the SNPs that flanked those that fell within a 1.5 LOD-drop of the top LOD score. We refer to these specific SNPs within these regions as those within the "1.5 LOD-drop interval". Top candidate SNPs were those that yielded the highest LOD scores in the SNP analysis and followed an expected founder haplotype pattern based on founder effects.

Candidate Gene Identification

Although all genes within the 95% Bayesian support interval of each bQTL were considered to be candidate genes, those that possess the most supporting evidence from other studies are considered to be those that are most likely the effective contributors to the QTL signal (i.e. "top candidate genes"). The list of candidates was further narrowed down to genes: 1) that contained significant cis-acting expression quantitative trait loci (cis-eQTL) in mouse liver tissue; 2) whose expression levels were significantly ethanol-responsive (or different between acute ethanol-treated and control mice) BXD mouse reward system-related brain regions; 3) correlated with ethanol-drinking-related phenotypes in chronically voluntarily drinking (under an IEA paradigm, with 22hsr of access to 4% EtOH and 2hrs of abstinence per day) male Rhesus Macaque anterior cingulate cortex and subgenual cortex (168, 169). Genes with the most supporting evidence from these studies were considered to be the top candidates; or 4) significantly associated with alcohol-related traits in any study reported in GWAS Catalog (170); or 5) harboring a significant weekly alcohol consumption-associated SNP the GWAS & Sequencing Consortium of Alcohol and Nicotine (GSCAN) (61). However, genes were not excluded as candidates for not having supporting evidence from any one of these studies.

Expression QTL data from liver tissue (the site of ethanol metabolism) was obtained from Munger et al. (171), which examined basal expression levels in the liver of DO mice (assessed via RNAseq), and assayed genotypes via the MUGA (the original version of the DO-catered arrays) for genotyping. Expression QTL with a $q_{Bootstrap} < 0.05$ were considered to be significant. Ethanol-responsiveness was defined as significant ($q_{FisherFDR} < 0.05$) different expression levels, for at least one probeset representing that gene, in the Nucleus

Accumbens, Prefrontal Cortex, or Ventral Tegmental Area between ethanol-treated (via intraperitoneal injection) mice and acute saline-treated mice, in a sample of mice from 27 BXD strains described in Wolen et al. (27). These three brain regions were chosen due to their role in the mesolimbocortical pathway known to be involved in alcohol craving (modeled by the post-abstinence increase in ethanol consumption in IEA mouse paradigms).

Genes were considered to be supported by evidence from Bogenpohl et al. (in preparation) if their expression values were suggestively (unadjusted $p<0.05$) correlated with any one of the following phenotypes: treatment group (alcohol-exposed vs. water-only); drinking category (scale of 0-4, from no alcohol consumption to high, binge-like levels of consumption); average blood ethanol content (BEC) (measured roughly every 5 days) across the whole study, the first six months of drinking, or the last six months of drinking; change in BEC between first and last six months; average daily ethanol intake across the entire study, first six months, and last six months; change in ethanol intake between first and last six months; or average daily ethanol preference (g/kg(EtOH)/mL(172)); or if they were significantly differentially expressed between ethanol-exposed monkeys and controls.

Results

Behavioral Measures

Daily total ethanol consumption values can be seen in Figure 17. As expected, each behavior varied quite widely across samples, with last week mean total ethanol consumption ranging from 0g/kg to 40g/kg (Fig. 17-18). A Log-transformation ($\log_{10}(1+\text{value})$) for ethanol consumption and 30% ethanol choice, and a square-root transformation for ethanol

preference, worked best for normalizing the respective phenotypic distributions before QTL analysis (Fig. 18). There was an overall significant increase in alcohol consumption between the first week and last week of IEA (Fig. 19). Although the whole sample's trend was to increase drinking over time, this was not the case for every individual mouse, as we predicted based on similar variation between the progenitor strains in our previous study *Chapter 3*. Phenotypes were moderately correlated across time points, with enough variation between them to justify analyzing them separately for QTL mapping (Fig. 20; Table 2).

Quantitative Trait Loci

Several significant and suggestive QTL were identified for each of the phenotypes measured (total ethanol consumption, total ethanol preference, and 30% ethanol choice), all of which differed between the first and last week experimental time ranges (Table 3, Fig. 21). Support intervals for these loci ranged from an unprecedented 1.05Mbp to 25.38Mbp. One of the two significant loci (Chr4:3.37-11.04). was identified for last week mean total ethanol consumption, and was suggestive for last week mean total ethanol preference, and whole study mean consumption and preference. One other locus was identified for both ethanol consumption and preference, but for the first week time interval, on Chr12. Founder effects were somewhat ambiguous, even after plotting standard errors, so several different patterns were considered for each locus (Fig. 22).

Table 2: Correlations Within Phenotypes Across Time Intervals

Phenotype	First Week v. Last Week	First Week v. Whole Study	Last Week v. Whole Study
Total EtoH Cons.	r=0.51, p< 2.2e-16	r=0.80, p< 2.2e-16	r=0.89, p< 2.2e-16
Total EtoH Pref.	r=0.47, p< 2.2e-16	r=0.47, p< 2.2e-16	r=0.51, p< 2.2e-16
30% EtOH Choice	r=0.43, p< 2.2e-16	r=0.74, p< 2.2e-16	r=0.80, p< 2.2e-16

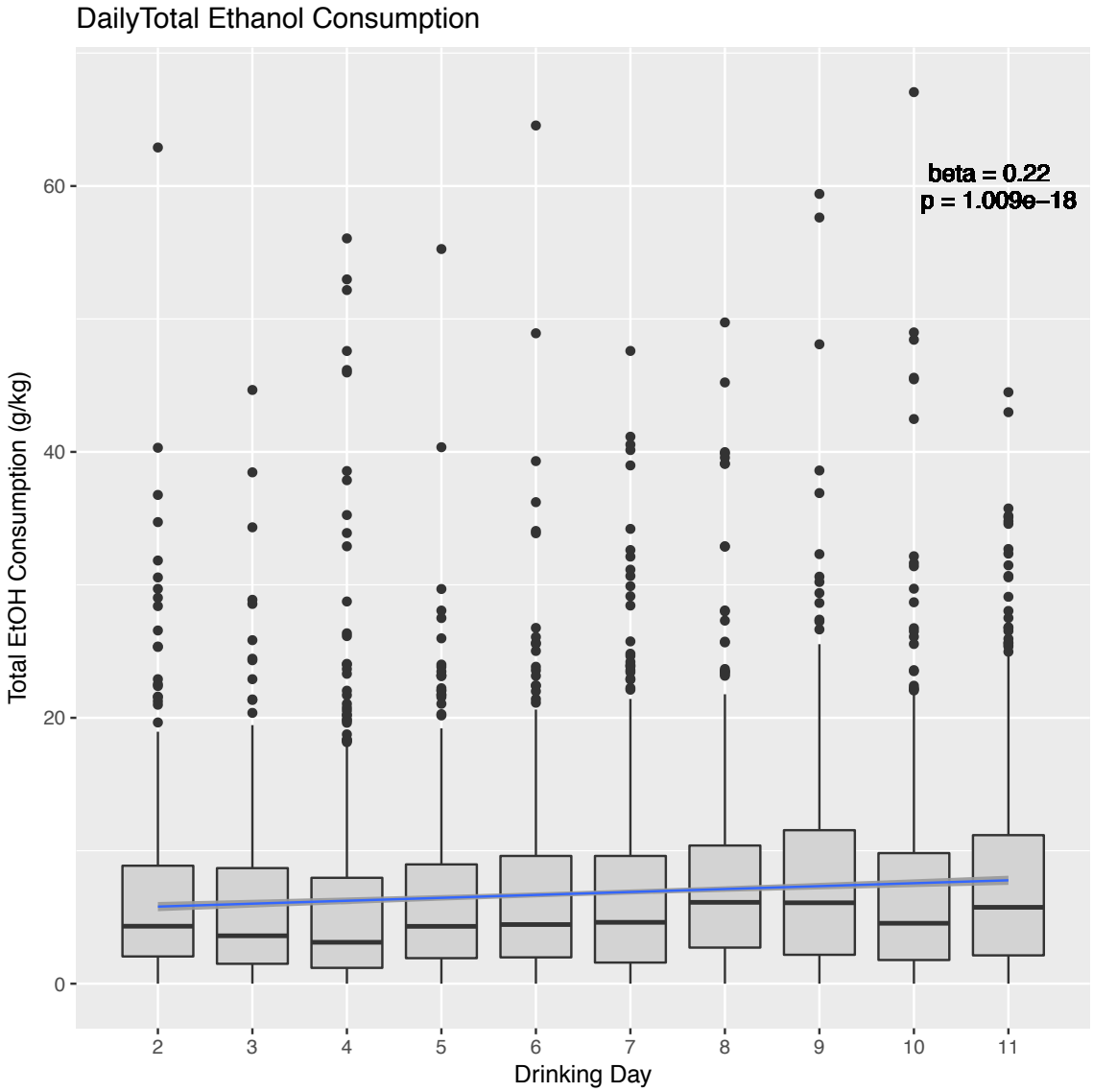
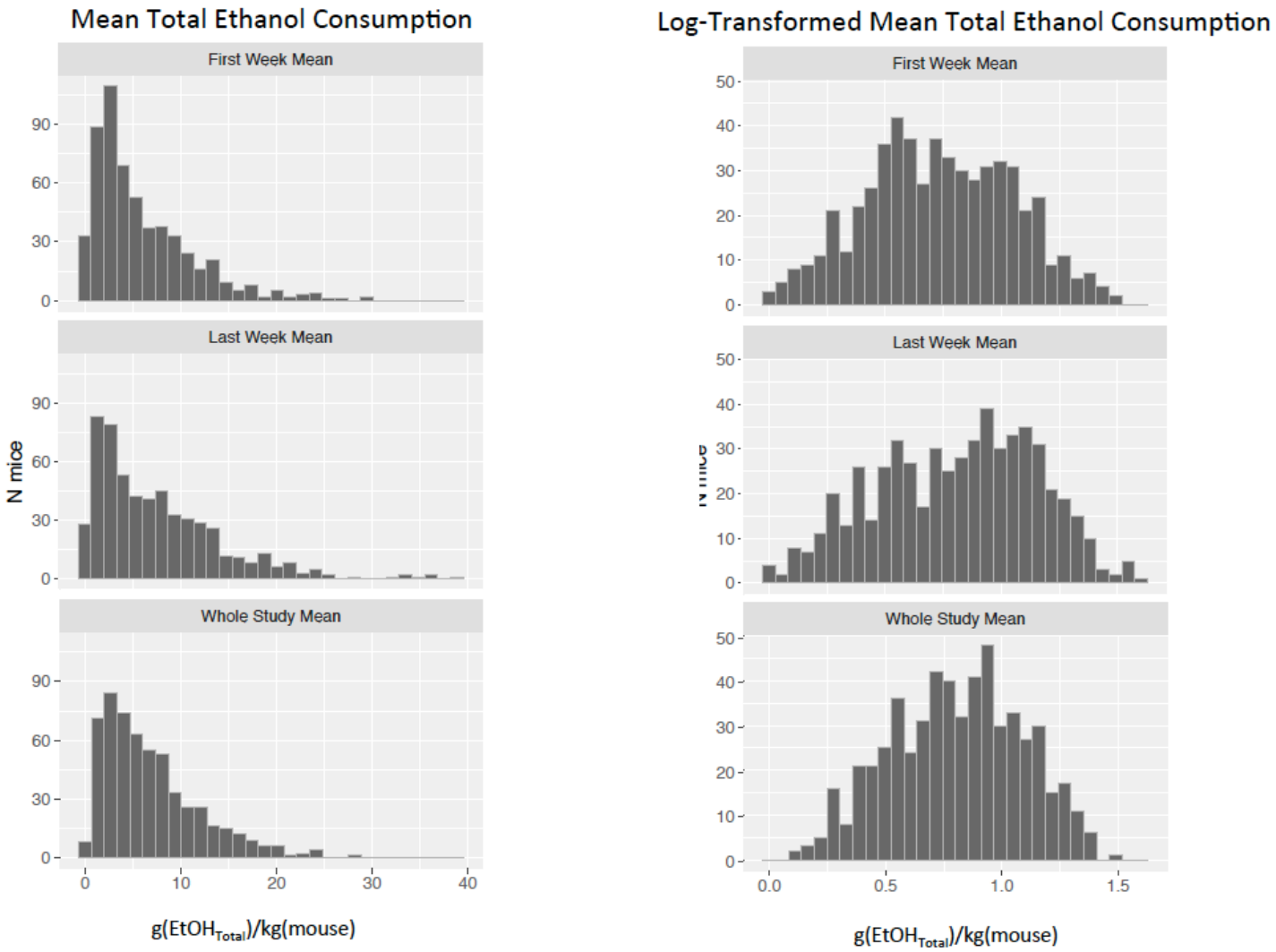
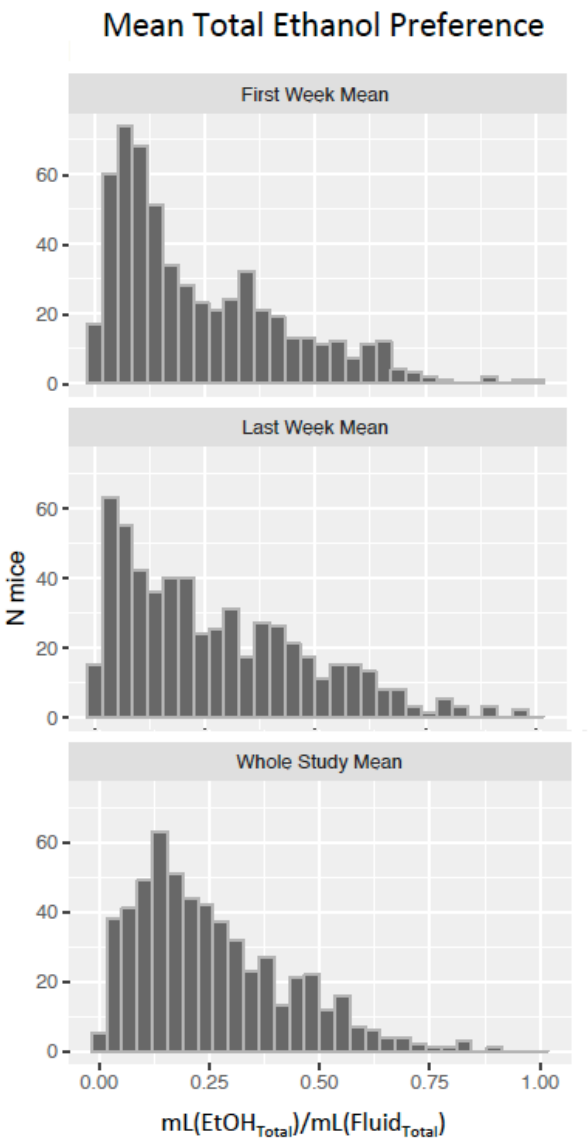


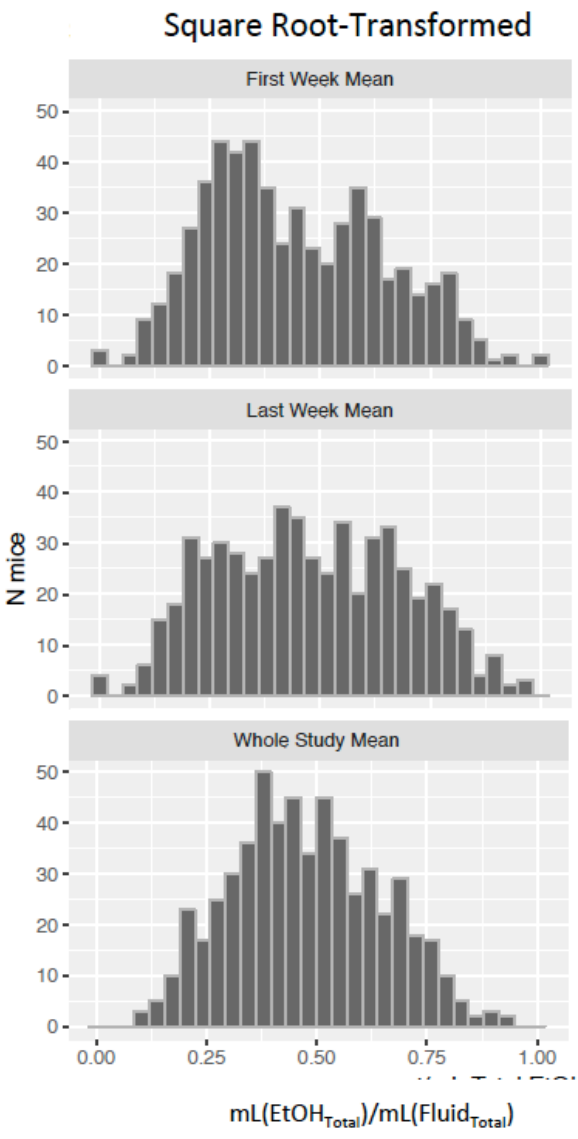
Figure 17: Slide 2: Daily total ethanol consumption (g(EtOH)/kg(mouse)) across the first 4 weeks of ethanol exposure, excluding the first day of exposure (ethanol access days 2-11), for all ethanol-exposed mice. Horizontal black lines represent the median, with grey blocks encapsulating the interquartile range, and vertical black lines extend across values calculated by subtracting (truncated on y-axis) or adding 1.5 multiplied by the interquartile range to the minimum or maximum values, respectively. Each black dot represents an individual mouse that fell outside of this range, which could be considered as potential outliers. The blue line is the line of best fit for the effect of drinking day on ethanol consumption, with the beta statistic and p-value for this regression depicted in the upper right-hand corner.

a



b





c

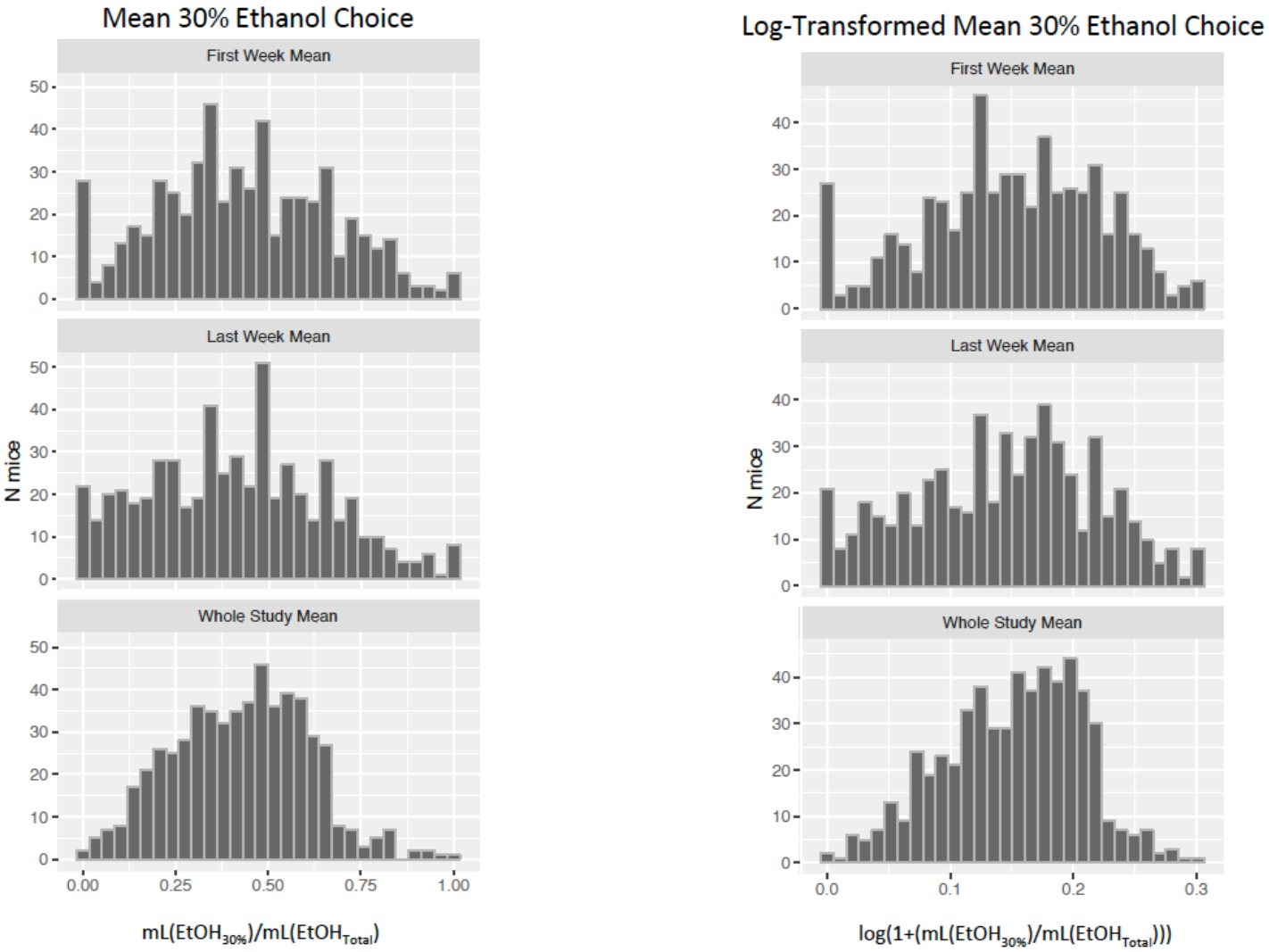


Figure 18: Histograms of total ethanol consumption (a), preference (b), and 30% ethanol choice (c) for the first week, last week, and entire duration of ethanol access, with number of mice exhibiting that level of consumption at each time point represented by the Y-axis. On the left are the raw values, and on the right are the log- or square-root-transformed values ($\log_{10}(1+x_{g/kg})$ or $\sqrt{x}$), which were used for QTL analysis due to the normal distribution assumption in the regression calculations.

SNP effects were analyzed by testing for a relationship between the phenotypes and imputed allelic probabilities at each individual locus (2 degrees of freedom, as opposed to 7), calculated from the haplotype probabilities (Fig. 23). Top-scoring SNPs are discussed in the context of their reported affiliated gene (one gene per SNP) by Sanger Mouse Genome (information that is imbedded in the R/qtl2 package), due to their close proximity. We chose to identify top candidate SNPs separately from top candidate genes, because it is unlikely that every SNP would only affect the gene to which it is most proximal. Candidate SNPs represent polymorphisms that are contributing to the signal, and candidate genes refer to the functional entities that may be affected by these SNPs. SNP analysis revealed LOD peaks in which many of the highest-scoring SNPs (within a 1.5 LOD-drop from the top-scoring SNP) matched the expected founder haplotype distribution. We will hence forth refer to SNPs with the 5 highest LOD scores, of those that match the expected founder haplotype patterns, as "candidate SNPs". For all but three loci (Chr3 Locus 3, the Chr11 locus for Whole Study Total Ethanol Consumption, and the Chr6 Locus), all of the highest scoring SNPs followed the expected founder haplotype patterns (Table S1). The SNPs with the top 5 LOD scores for each locus are listed in Table 4. However, the top SNPs for the Chr6 locus were shared between NOD, NZO, and CAST, and a nearby SNP contained a C57-unique allele. These alleles are likely both contributing to the founder effects, together. This could explain the higher mean drinking values in mice possessing haplotypes inherited from any one of these 5 founder strains. Candidate genes were identified as those in which the SNPs were contained.

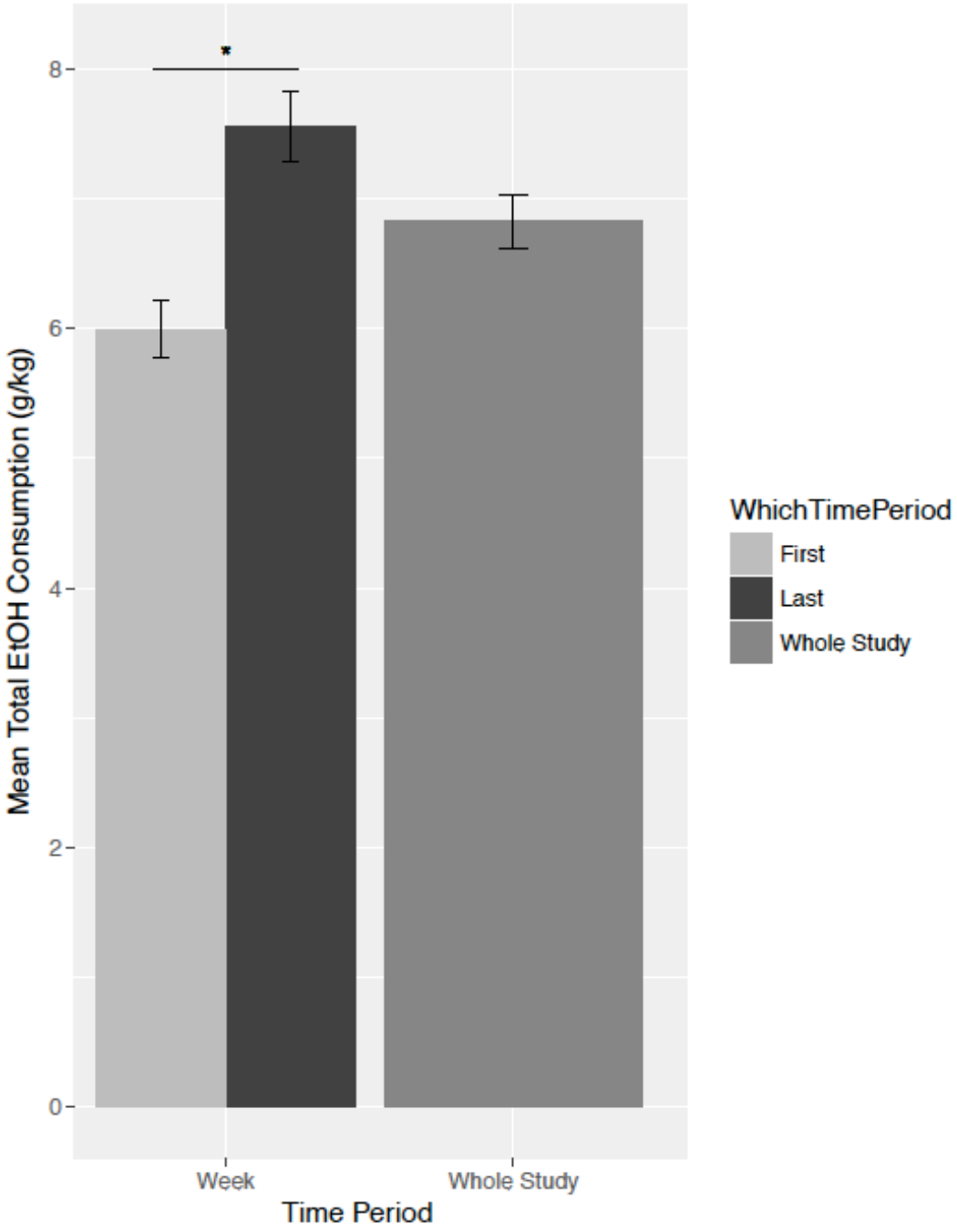


Figure 19: Grand mean and standard errors (vertical black lines) of total ethanol consumption (g(EtOH)/kg(mouse)) across all mice, for the first week, last week, and entire duration of ethanol exposure. The bar and asterisk indicate a significant difference (p<0.05) between total ethanol consumption between those time intervals.

Top Candidate Genes

All bQTL with significant or suggestive loci were examined for top candidate genes. As described above, these genes were assessed for: significant eQTL in DO liver tissue; suggestive (unadjusted $p<0.05$) anterior cingulate cortex and subgenual cortex gene expression correlation with ethanol-consumption-related trait in Rhesus Macaques; significant expression differences between acute ethanol treatment and control BXD mouse reward pathway-related brain regions; and significant GWAS signals for alcohol-related-traits. A summary of supporting evidence for each of the top candidate genes can be seen in Table 5. All supporting evidence for all genes within each bQTL's support intervals can be seen in Table S2-3.

Although the support intervals differed slightly between phenotypes, the Chromosome 4 QTL was identified for *Whole Study Mean Total Ethanol Preference* and *Consumption*, as well as *Last Week Mean Total Ethanol Preference* and *Consumption*. The top candidate genes for this region are: *Car8* (Carbonic anhydrase 8), *Fam110b* (Family with sequence similarity 110, member B), *Impad1* (Inositol monophosphatase domain-containing 1), and *Tox* (Thymocyte selection associated high mobility group box). Because of its narrow support interval, the *Last Week Mean Total Ethanol Consumption* QTL only contained one of the top candidate genes (*Car8*), whereas all other regions shared all candidate genes for this region. The top candidate SNPs for this QTL were determined by pooling results from each of these analyses. These SNPs are: rs32098535, rs48821381, rs227159729, rs27657700, and rs51487109 (Table 4). Two of these SNPs are not known to be associated with any genes, and three others are non-coding transcript variants in *Tox*.

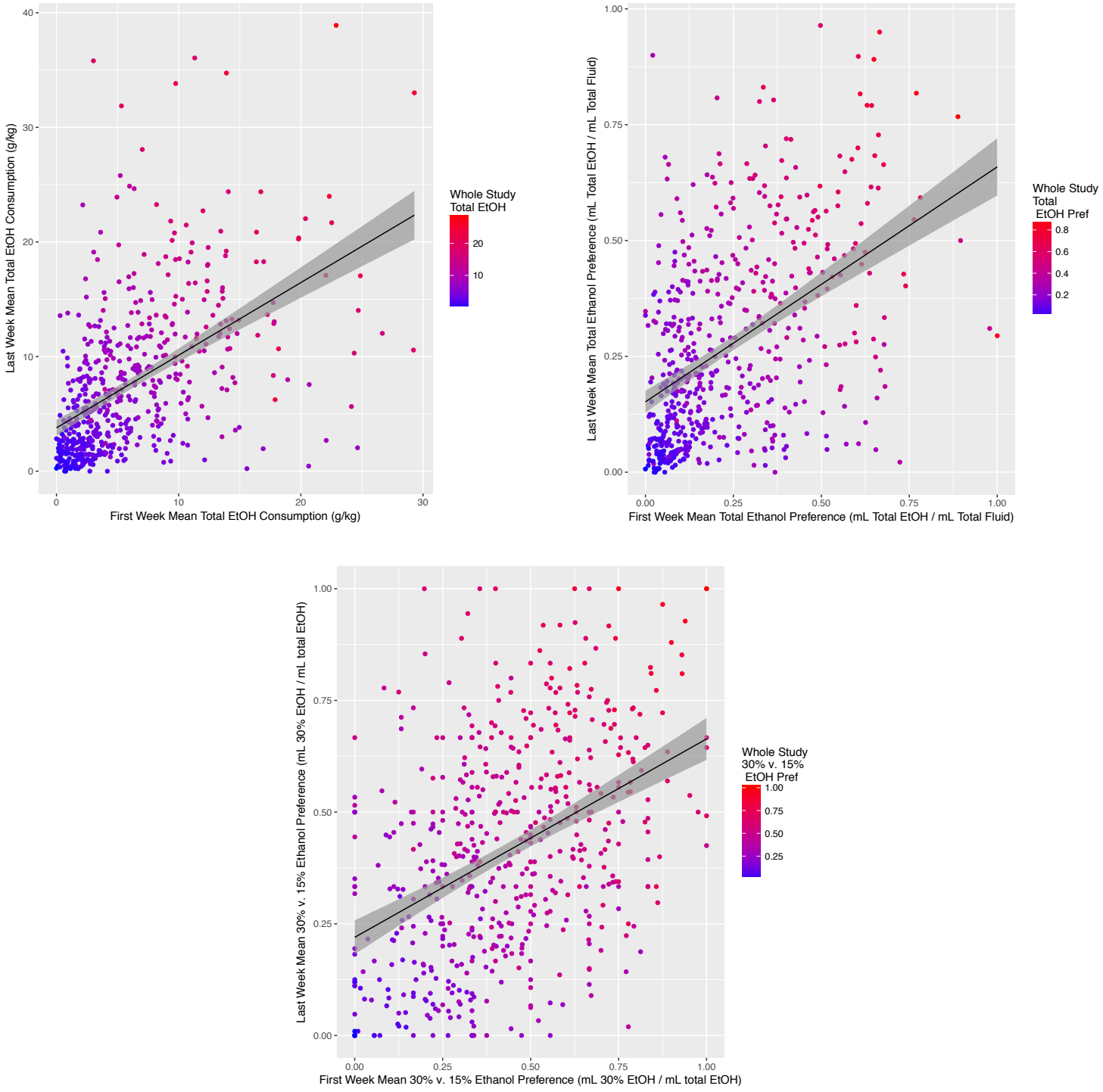


Figure 20: Relationship of average phenotypes (a. total ethanol consumption, b. total ethanol preference, c. 30% ethanol choice) between time points (First Week, Last Week, and Whole Study). Each point represents a single mouse, and its color reflects the Whole Study mean phenotype. The black line represents the line of best fit between the First and Last Week mean phenotypes, with the shaded grey area representing the standard error.

Table 3: bQTL Regions for Ethanol Drinking Phenotypes

Phenotype	marker	LOD Score	Chr	pos_Mbp	Downstream 95% Bayesian CI	Upstream 95% Bayesian CI	CI Length
Last Week Mean 30% EtOH Choice	JAX00221237	8.63	3	108.2251	105.9240	109.6894	3.7654
Last Week Mean Total EtOH Cons.	UNC6694526	8.23	4	8.4134	8.2352	9.2852	1.0500
First Week Mean Total EtOH Pref.	UNC21463670	7.52	12	79.3497	78.0891	79.9077	1.8185
Last Week Mean 30% EtOH Choice	UNCHS009636	6.93	3	103.5692	101.8533	104.4704	2.6171
Last Week Mean Total EtOH Pref.	UNC28339832	6.88	17	74.3198	72.9064	75.2112	2.3048
Whole Study Mean Total EtOH Cons.	UNCHS010637	6.67	4	8.3241	3.7755	9.6532	5.8776
Whole Study Mean Total EtOH Pref.	UNC1732354	6.65	1	135.3548	134.6418	136.5865	1.9448
Last Week Mean 30% EtOH Choice	UNC6068240	6.48	3	118.9298	118.3936	119.8124	1.4189
Last Week Mean Total EtOH Cons.	c11.loc26.06	6.46	11	50.5751	49.9637	51.5240	1.5603
Last Week Mean Total EtOH Pref.	UNCHS042680	6.36	16	44.6870	39.4349	45.6827	6.2479
Last Week Mean Total EtOH Pref.	UNCHS010637	6.32	4	8.3241	3.3708	11.0462	7.6754
Whole Study Mean Total EtOH Pref.	UNCHS010623	6.30	4	6.9963	3.3708	8.3241	4.9533
Whole Study Mean Total EtOH Cons.	UNCHS031224	6.29	11	70.6117	69.8561	71.3399	1.4838
First Week Mean Total EtOH Cons.	c6.loc49.42	6.27	6	113.9204	112.7439	115.3086	2.5647
Last Week Mean Total EtOH Pref.	UNCHS002448	6.06	1	134.7291	119.5259	144.9080	25.3821

Behavioral quantitative trait loci positions for Last Week, First Week, and Whole Study Mean Total Ethanol Consumption and Preference, and 30% Ethanol Choice. Peak LOD scores are annotates with their respective genotyping array markers, and their target position in Mega base pairs (Mbp) on the indicated chromosome, for each QTL region. 95% Bayesian confidence intervals (CI) were used to determine QTL length.

As for the Chr4 QTL, a bQTL associated with a phenotype for more than one time interval (*Whole Study* and *Last Week Mean Total Ethanol Preference*) appeared on Chr1, with slightly different yet highly overlapping support intervals. The top candidate genes for both the *Whole Study* and *Last Week* loci are: *Csrp1* (Cysteine and glycine-rich protein 1), *Ipo9* (Importin 9), *Lad1* (Ladinin 1), *Lmod1* (Leiomodin 1), *Nav1* (Neuron navigator 1), *Rnpep* (Arginyl aminopeptidase), *Tnni1* (Troponin I, skeletal slow 1), *Tnnt2* (Troponin T2, cardiac). The top candidate genes unique to the larger *Last Week* locus are: *Cd55* (CD55 molecule, decay accelerating factor for complement), *Cd55b* (CD55 molecule, decay accelerating factor for complement B), *Nucks1* (nuclear casein kinase and cyclin-dependent kinase substrate 1), *Pm20d1* (peptidase M20 domain containing 1), *Slc41a1* (solute carrier family 41, member 1), and *Timm17a* (translocase of inner mitochondrial membrane 17a). The top five candidate SNPs in the variant-based analysis were rs32438212, rs51581531,

rs48738822, rs108288159, and rs46404195, all of which followed the expected founder haplotype patterns and had LOD scores that rounded to 7.36 (Table 4).

Three QTL were identified on Chr3 for Last Week Mean 30% Ethanol Choice. However, the support intervals of the two most proximal loci were only 1.5Mbp apart, and only one distinct LOD peak is seen in this region (Fig. 22), indicating that these two peaks are actually identifying one locus that spans over 7.84Mbp. The top candidate genes for these two loci are: *Ahcyl1* (S-adenosylhomocysteine hydrolase-like 1), *Ampd2* (adenosine monophosphate deaminase 2), *Cd53* (Cd53 antigen), *Clcc1* (chloride channel clic-like 1), *Gnai3* (guanine nucleotide binding protein, alpha inhibiting 3), *Sort1* (Sortilin 1), *Ubl4b* (ubiquitin-like 4b), *Wdr77* (WD repeat domain 77), *Hipk1* (homeodomain interacting protein kinase 1), *Rsbn1* (rosbin, round spermatid baseic protein 1), *Syt6* (synaptotagmin 6), and *Tspan2* (tetraspanin 2). Genes that fell between these two support intervals were also considered, but none displayed substantial evidence for being a top candidate gene. The third locus, which appears to be a distinct peak distal to the other two (Fig. 22), was a narrower region containing only two top candidate genes: *Dpyd* (dihydropyrimidine dehydrogenase) and *Ptbp2* (polypyrimidine tract binding protein 2). The top candidate SNPs for the most proximal locus are rs49087152, rs217960262, rs244300971, rs246339091, and rs213764891 (Table 4), which all have LOD scores between 6.06 and 6.07.

Two loci were identified for *Last Week Mean Total Ethanol Preference*, on Chromosomes 16 and Chr17. The top candidate genes for the Chr16 locus are: *Lsamp* (limbic system-associated membrane protein), *Slc35a5* (solute carrier family 35, member A5), and *Zbtb20* (zinc finger and BTB domain containing 20). The top five candidate SNPs for the Chr16 locus are rs243048813, rs32655787, rs257335011, and rs47507732, all of which are

intron variants for genes *Sidt1* (SID1 transmembrane family, member 1) (for rs243048813) and *Lsamp* (for the remaining four SNPs) with LOD scores between 3.87 and 4.46 (Table 4). For Chr17, the top candidate genes are: *Capn13* (calpain13), *Dpy30* (dpy-30, histone methyltransferase complex regulatory subunit), *Galnt14* (polypeptide N-acetylgalactosaminyltransferase 14), *Lbh* (limb-bud and heart), *Lclat1* (lysocardiolipin acyltransferase), *Memo1* (mediator of cell motility 1), *Spast* (spastin), and *Yipf4* (yip 1 domain family, member 4). The top five candidate SNPs for the Chr17 locus are rs51050012, rs251169915, rs51169739, rs222840432, and rs235513168, which have LOD scores between 3.99 and 4.01. Two of these SNPs (rs51169739 and rs222840432) are intron variants for the gene *Lclat1* (lysocardiolipin acyltransferase 1), one is an downstream variant of predicted gene *Gm25406*, and the remaining two are classified as intergenic SNPs.

There were two loci that were unique to ethanol-naïve (i.e. *First Week*) drinking behaviors: one on Chr6 for *Total Ethanol Consumption*, and the other on Chr12 for *Total Ethanol Preference*. The top candidate genes for the Chr6 locus are: *Atp2b2* (ATPase, Ca^{2+} transporting, plasma membrane 2), *Slc6a1* (solute carrier family 6, member 1), and *Slc6a11* (solute carrier family 6, member 11). The top five candidate SNPs for this region are rs30214990, rs46370242, rs252343765, rs36751483, rs48409474, which have LOD scores between 3.19 and 3.36. All of these SNPs are all located within the *Atp2b2* gene. The Chr12 top candidate genes are: *Arg2* (arginase type II), *Gphn* (gephyrin), *Mpp5* (membrane protein, palmitoylate 5), *Plekhh1* (pleckstrin homology domain containing, family H (with MyTH4 domain), member 1), and *Zfyve26* (zinc finger, FYVE domain containing 26). The top five candidate SNPs for this locus were rs226566367, rs46944281, rs48897594, rs48645692, and rs233426717, which fell within the introns or just up-or down-stream of genes *Zfyve26*,

Rdh12 (retinol dehydrogenase 12), *Rdh11* (retinol dehydrogenase 11), *Arg2*, and *Plekhh1*, respectively. There was also one locus unique to a *Whole Study* ethanol behavior, on Chr11, for *Total Ethanol Consumption*, and an adjacent but distinct locus unique to *Last Week Mean Total Ethanol Consumption*. The former contained several top candidate genes: *0610010K14Rik* (Riken cDNA 0610010K14), *Alox12* (arachidonate 12-lipoxygenase), *Arrb2* (arrestin, beta 2), *Camta2* (calmodulin binding transcription activator2), *Med11* (mediator complex subunit 11), *Mink1* (misshapen-like kinas 1), *Pelp1* (proline, glutamic acid and leucine rich protein1), *Phf23* (PHD finger protein 23), *Psmb6* (proteasome), *Rabep1* (rabaptin, RAB GTPase binding effector protein 1), RNAsek (ribonuclease RNase K), and *Tm4sf5* (transmembrane superfamily 5). The top five candidate SNPs are 11:70781749_C/T, 11:70779940_A/T, 11:70618113_C/A, 11:70782944_T/G, and 11:70611698_A/G, each with a LOD score of 1.30. Unexpectedly, the top-scoring SNPs, with LODs of 2.74, did not match the expected founder haplotype patterns. Therefore, we have chosen not to rely on SNP information to identify potential affective genes for this locus. Finally, the *Last Week Mean Total Ethanol Consumption* QTL on Chr11 contained three top candidate genes: *Canx* (calnexin), *Clk4* (CDC like kinase 4), and *Zfp2* (zinc finger protein 2). The top candidate SNPs for this locus are rs258127079, rs243374395, rs26965630, rs261774330, and rs26979779. Two of these (rs26965630 and rs26979779) SNPs are not contained in any genes. However, the other 3 are classified as intron (rs258127079 and rs261774330) or upstream (rs243374395) variants for the gene *Adamts2*.

Table 4: Top-Scoring Variants within QTL Support Intervals

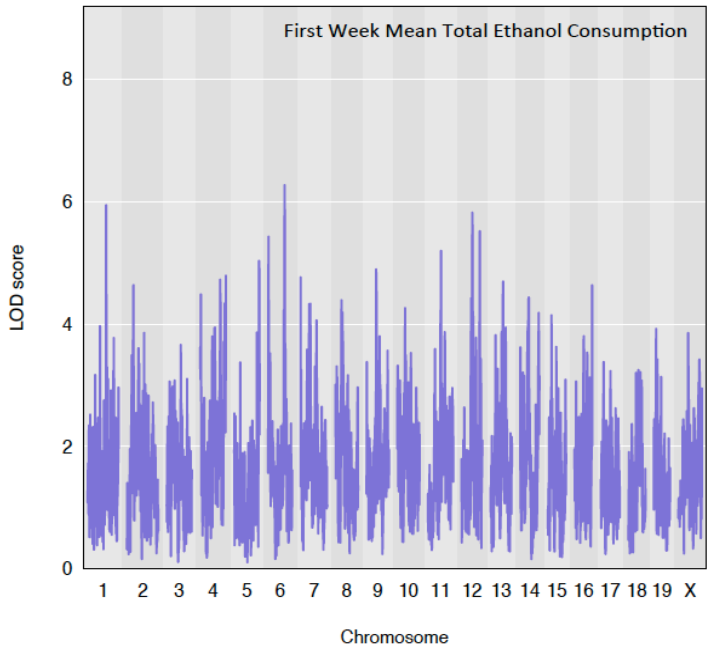
Phenotype	Which Locus	LOD Score	Pos (Mbp)	SNPid	Gene	Consequence of Minor Allele	Variant Type
Whole Study Total EtOH Cons	Chr4 Locus	3.05	8.3842	4:8384190_G/A	NA	intergenic_variant	snp
		3.02	8.4380	rs249655952	Gm37386	downstream_gene_variant	snp
		2.89	8.3023	4:8302313_A/C	NA	intergenic_variant	snp
		2.75	7.9466	rs27708815	NA	intergenic_variant	snp
		2.38	7.8969	rs27655790	NA	intergenic_variant	snp
Last Week Total EtOH Cons	Chr4 Locus	4.01	8.4380	rs249655952	Gm37386	downstream_gene_variant	snp
		3.89	8.3842	4:8384190_G/A	NA	intergenic_variant	snp
		3.43	8.3023	4:8302313_A/C	NA	intergenic_variant	snp
		3.15	9.1646	rs32455519	NA	intergenic_variant	snp
		3.10	9.2412	rs262840752	NA	intergenic_variant	snp
Whole Study Total Ethanol Preference	Chr4 Locus	3.70	7.2540	rs32098535	NA	intergenic_variant	snp
		3.62	7.5980	rs48821381	NA	intergenic_variant	snp
		3.62	7.5982	rs27687625	NA	intergenic_variant	snp
		3.13	6.7412	rs227159729	Tox	non_coding_transcript_exon_variant,non_coding_transcript_variant	snp
		3.12	6.7416	rs27657700	Tox	non_coding_transcript_exon_variant,non_coding_transcript_variant	snp
Last Week Total Ethanol Preference	Chr4 Locus	3.74	5.2230	SV_4_5223003_5223005	NA	NA	SV
		3.74	5.2577	SV_4_5257666_5257668	NA	NA	SV
		3.74	6.6281	SV_4_6628120_6628122	NA	NA	SV
Whole Study Total Ethanol Preference	Chr1 Locus	3.74	135.7212	rs32438212	Csrp1	intron_variant	snp
		3.74	135.7238	rs51581531	Csrp1	intron_variant	snp
		3.74	135.7263	rs45936880	Csrp1	intron_variant	snp
		3.74	135.7277	rs45881802	Csrp1	intron_variant	snp
		3.74	135.7311	rs46994917	Csrp1	intron_variant	snp
Last Week Total Ethanol Preference	Chr1 Locus	3.85	135.3568	rs581908485	Lmod1	intron_variant	snp
		3.80	130.4435	1:130443512_G/C	Cd55	intron_variant	snp
		3.80	130.4436	1:130443642_A/G	Cd55	intron_variant	snp
		3.80	130.4437	1:130443663_A/C	Cd55	intron_variant	snp
		3.80	130.4438	1:130443779_G/T	Cd55	intron_variant	snp
Whole Study Total EtOH Cons	Chr11 Locus	2.72	70.1879	11:70187893_G/A	NA	intergenic_variant	snp
		2.72	70.1879	11:70187943_G/A	NA	intergenic_variant	snp
		2.72	70.1879	11:70187945_A/G	NA	intergenic_variant	snp
		2.72	70.1879	11:70187948_C/T	NA	intergenic_variant	snp
		2.72	70.1880	11:70187951_C/T	NA	intergenic_variant	snp
Last Week Total EtOH Cons	Chr11 Locus	4.02	50.6592	rs258127079	Adamts2	intron_variant	snp
		4.01	50.5990	rs243374395	Adamts2	upstream_gene_variant	snp
		4.00	50.5895	rs26965630	NA	intergenic_variant	snp
		4.00	50.5903	rs248344810	NA	intergenic_variant	snp
		4.00	50.5907	rs222454793	NA	intergenic_variant	snp
First Week Total EtOH Cons	Chr6 Locus	3.36	113.8618	rs30214990	Atp2b2	intron_variant,non_coding_transcript_variant	snp
		3.25	113.8328	rs46370242	Atp2b2	intron_variant	snp
		3.25	113.8250	rs252343765	Atp2b2	intron_variant	snp
		3.25	113.5486	rs36751483	Fancd2	upstream_gene_variant	snp
		3.20	113.9833	rs48409474	Atp2b2	intron_variant,non_coding_transcript_variant	snp
First Week Total Ethanol Preference	Chr12 Locus	6.14	79.2442	rs226566367	Zfyve26	intron_variant	snp
		6.14	79.2442	rs46652565	Zfyve26	intron_variant	snp
		6.14	79.2462	rs220212213	Zfyve26	splice_region_variant,intron_variant	snp
		6.14	79.2610	rs258744940	Zfyve26	intron_variant	snp
		6.14	79.2610	rs221141341	Zfyve26	intron_variant	snp
Last Week Total Ethanol Preference	Chr16 Locus	4.46	44.2681	rs243048813	Sidt1	intron_variant	snp
		4.46	44.2781	rs219475321	Sidt1	downstream_gene_variant	snp
		4.46	44.2781	rs257457254	Sidt1	downstream_gene_variant	snp
		4.46	44.2808	rs226215307	Sidt1	downstream_gene_variant	snp
		3.98	40.7403	rs32655787	Lsamp	intron_variant	snp
	Chr17 Locus	4.01	73.0464	rs51050012	NA	intergenic_variant	snp
		4.01	73.0467	rs33454915	NA	intergenic_variant	snp
		4.01	73.0474	rs29521103	NA	intergenic_variant	snp
		4.01	73.0486	rs33687405	NA	intergenic_variant	snp
		4.01	73.0522	rs33457925	NA	intergenic_variant	snp
Last Week Mean 30% EtOH Choice	Chr3 Locus1-2	6.07	108.7610	rs49087152	Aknad1	intron_variant,NMD_transcript_variant	snp
		6.07	108.7611	rs257383845	Aknad1	intron_variant,NMD_transcript_variant	snp
		6.07	108.7615	rs238046046	Aknad1	intron_variant,NMD_transcript_variant	snp
		6.07	108.7618	rs226352210	Aknad1	intron_variant,NMD_transcript_variant	snp
		6.07	108.7619	rs255653661	Aknad1	intron_variant,NMD_transcript_variant	snp

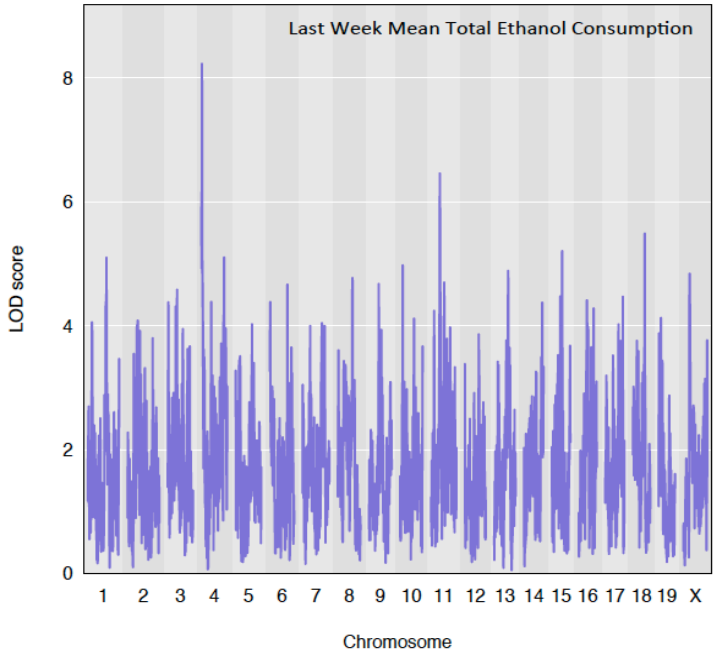
LOD scores, chromosomal position (in Mega base pairs (Mbp)), variant classification, SNP ID, and minor allele consequence for each of the top 5 variants within each QTL interval. Top variants were selected based on their LOD score and their adherence to the expected genotypic comparisons in the DO founder strains. If 5 SNPs did not meet the latter criterion, only those that did were selected.

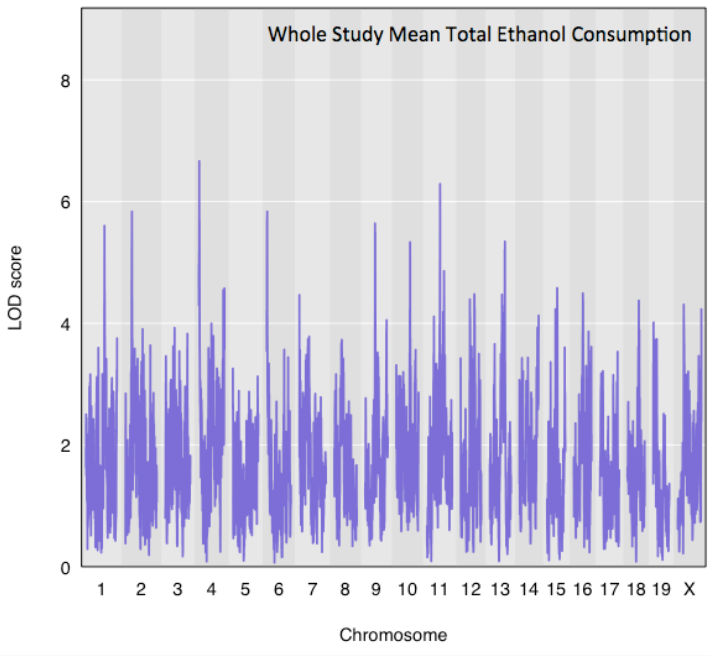
Table 5: Supporting Data for Top Candidate Variants

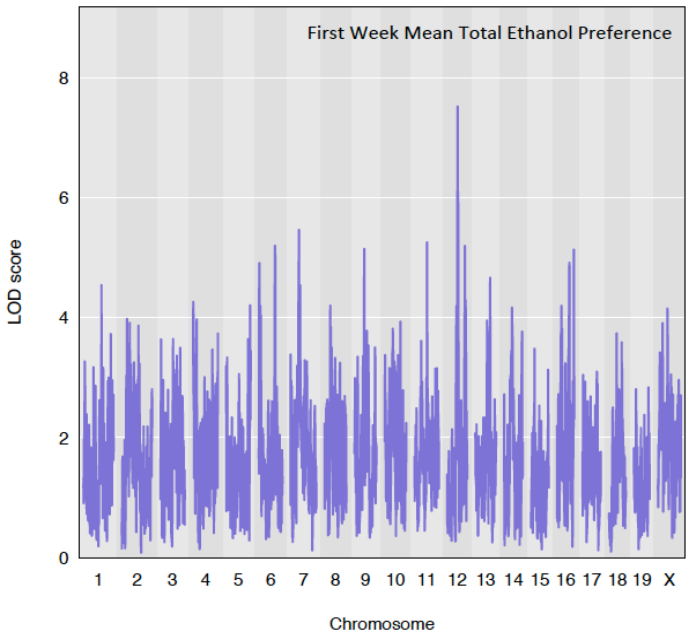
Gene	chr	start	stop	HumanGene	At Least 1 Sig. Value in Rhesus Data	Sig. cis-eQTL in DO Liver Data	N Sig. Alc. GWAS	Sig. S-score Nac	Sig. S-score PFC	Sig. S-Score VTA	Sig. in GSCAN
0610010K14Rik	11	70.2352	70.2379	NA	N	Y	0	N	Y	N	N
Ahcyl1	3	107.6631	107.6966	AHCYL1	Y	Y	0	Y	Y	Y	N
Alox12	11	70.2415	70.2554	ALOX12	Y	Y	0	N	N	N	N
Ampd2	3	108.0741	108.0867	AMPD2	N	NA	0	Y	Y	N	N
Arg2	12	79.1308	79.1563	ARG2	Y	Y	0	Y	N	N	N
Arrb2	11	70.4326	70.4408	ARRB2	N	Y	0	Y	Y	Y	N
Atp2b4	1	133.6995	133.8010	ATP2B4	Y	Y	0	N	N	N	N
Camta2	11	70.6695	70.6882	CAMTA2	Y	Y	0	N	Y	Y	N
Capn13	17	73.3065	73.3993	CAPN13	Y	NA	0	N	N	N	N
Car8	4	8.1415	8.2390	CA8	Y	Y	0	N	Y	Y	N
Cd53	3	106.7589	106.7901	CD53	Y	Y	0	N	Y	N	N
Cd55	1	130.4298	130.4630	CD55	Y	Y	0	N	N	N	N
Cd55b	1	130.3885	130.4230	NA	Y	NA	0	N	N	N	N
Chd7	4	8.6904	8.8684	CHD7	Y	NA	0	Y	N	Y	N
Clcc1	3	108.6539	108.6788	CLCC1	Y	Y	0	Y	Y	N	N
Csrp1	1	135.7201	135.7522	CSRP1	Y	Y	0	Y	Y	Y	N
Dpy30	17	74.2995	74.3239	DPY30	Y	Y	0	N	N	N	N
Dpyd	3	118.5621	119.4329	DPYD	Y	Y	0	N	N	N	N
Fam110b	4	5.6441	6.1082	FAM110B	Y	Y	0	N	N	N	N
Galnt14	17	73.4932	73.7108	GALNT14	N	Y	0	N	N	N	N
Gphn	12	78.2261	78.6848	GPHN	Y	Y	0	Y	Y	Y	N
Hipk1	3	103.7398	103.7919	HIPK1	Y	Y	0	Y	Y	Y	N
Impad1	4	4.7625	4.7934	IMPAD1	Y	Y	1	Y	Y	Y	N
Ipo9	1	135.3823	135.4305	IPO9	Y	Y	0	N	N	N	N
Lad1	1	135.8186	135.8333	LAD1	Y	Y	0	N	N	N	N
Lbh	17	72.9183	72.9419	LBH	Y	NA	0	Y	Y	Y	N
Lclat1	17	73.1080	73.2434	LCLAT1	Y	Y	0	N	N	N	N
Lmod1	1	135.3248	135.3681	LMOD1	Y	NA	0	N	N	N	N
Lsamp	16	39.9844	42.1817	LSAMP	N	Y	2	Y	Y	Y	N
Med11	11	70.4519	70.4537	MED11	Y	Y	0	N	N	N	N
Memo1	17	74.1990	74.2955	MEMO1	N	Y	0	Y	N	N	N
Mink1	11	70.5627	70.6145	MINK1	Y	Y	0	Y	N	Y	N
Mpp5	12	78.7489	78.8407	MPP5	N	Y	0	Y	Y	Y	N
Nav1	1	135.4346	135.6881	NAV1	N	N	0	Y	Y	Y	N
Nucks1	1	131.9105	131.9363	NUCKS1	Y	Y	0	Y	Y	Y	Y
Pelp1	11	70.3929	70.4100	PELP1	N	Y	0	Y	N	Y	N
Phf23	11	69.9958	70.0000	PHF23	Y	NA	0	N	N	N	N
Plekhh1	12	79.0292	79.0817	PLEKHH1	Y	Y	0	Y	N	Y	N
Pm20d1	1	131.7974	131.8215	PM20D1	N	Y	0	N	N	N	Y
Psmb6	11	70.5254	70.5279	PSMB6	Y	Y	0	N	N	N	N
Ptbp2	3	119.7187	119.7845	PTBP2	Y	NA	0	Y	Y	Y	N
Rab2a	4	8.5356	8.6078	RAB2A	Y	Y	0	N	N	N	N
Rabep1	11	70.8447	70.9431	RABEP1	Y	Y	0	N	N	Y	N
Rnasek	11	70.2381	70.2399	RNASEK	N	Y	0	N	N	Y	N
Rnpep	1	135.2627	135.2844	RNPEP	N	Y	0	Y	N	Y	N
Rsbn1	3	103.9141	103.9666	RSBN1	Y	NA	0	N	Y	Y	N
Slc16a1	3	104.6387	104.6585	SLC16A1	Y	NA	0	N	N	N	N
Slc16a11	11	70.2128	70.2164	SLC16A11	N	Y	0	N	N	N	N
Slc35a5	16	45.1396	45.1587	SLC35A5	Y	Y	0	N	N	Y	N
Slc41a1	1	131.8275	131.8489	SLC41A1	Y	Y	0	N	N	Y	Y
Sort1	3	108.2841	108.3615	SORT1	Y	Y	0	Y	Y	Y	N
Spast	17	74.3375	74.3911	SPAST	Y	Y	0	N	N	N	N
Syt6	3	103.5752	103.6456	SYT6	N	Y	0	Y	N	Y	N
Timm17a	1	135.2952	135.3138	TIMM17A	Y	Y	0	N	N	N	N
Tm4sf5	11	70.5052	70.5112	TM4SF5	N	NA	1	N	N	N	N
Tnni1	1	135.7794	135.8110	TNNI1	Y	NA	0	N	N	Y	N
Tnnt2	1	135.8363	135.8523	TNNT2	Y	NA	0	N	N	N	N
Tox	4	6.6864	6.9916	TOX	Y	Y	0	Y	Y	N	N
Tspan2	3	102.7345	102.8015	TSPAN2	Y	N	0	Y	Y	Y	N
Ubl4b	3	107.5537	107.5551	UBL4B	Y	NA	1	N	N	N	N
Wdr77	3	105.9594	105.9700	WDR77	Y	Y	0	N	N	Y	N
Yipf4	17	74.4895	74.5003	YIPF4	Y	Y	0	Y	N	Y	N
Zbtb20	16	42.8758	43.6426	ZBTB20	Y	Y	0	Y	Y	Y	N
Zfyve26	12	79.2323	79.2963	ZFYVE26	Y	Y	0	Y	N	N	N

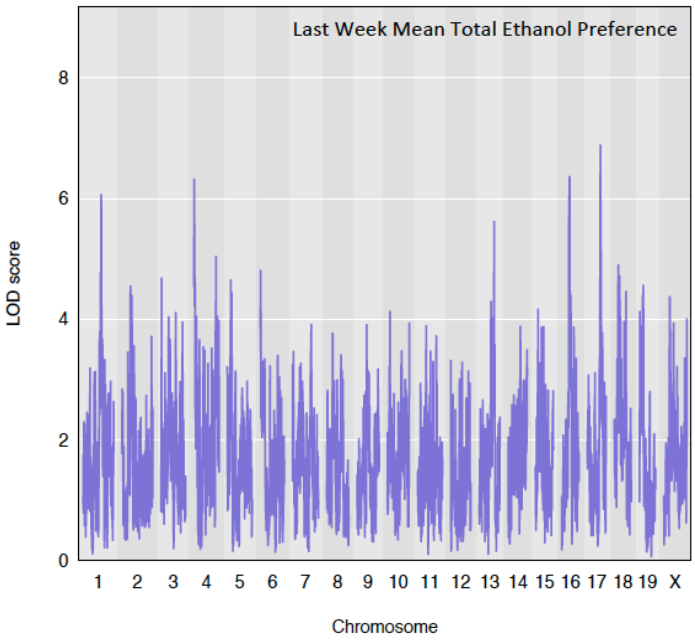
Top candidate genes, and their chromosomal locations in Mega base pairs, and a summary of supporting data for association with alcohol-related phenotypes, gleaned from other studies.
N=No, Y=Yes

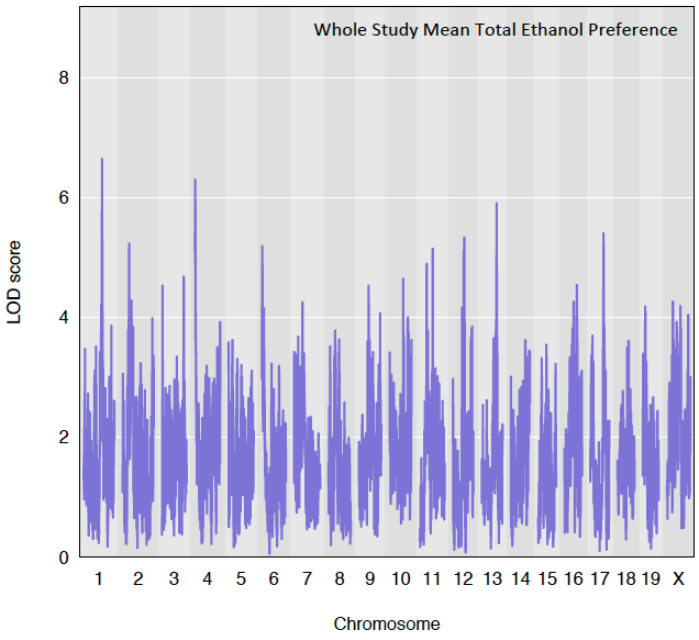












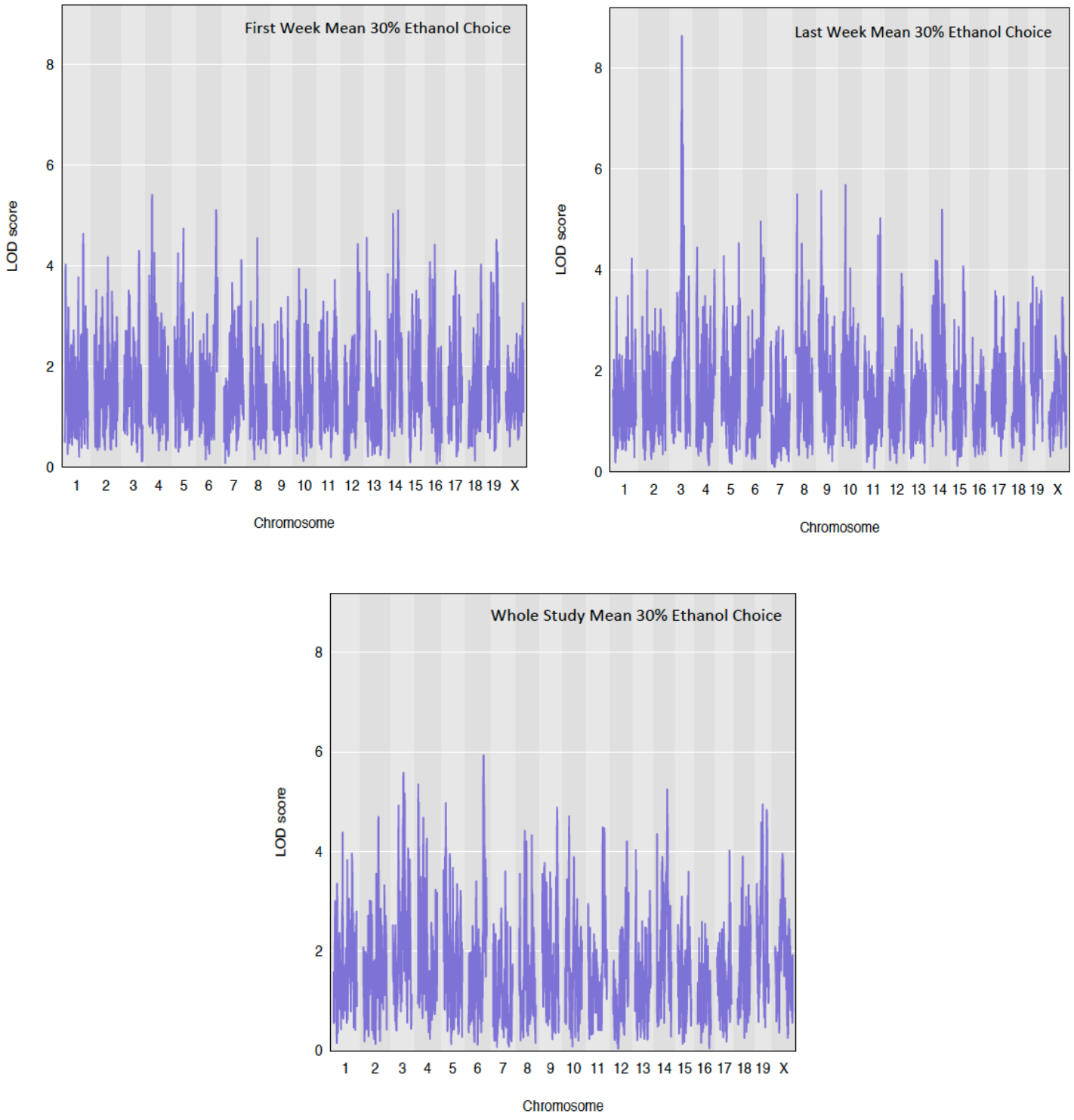
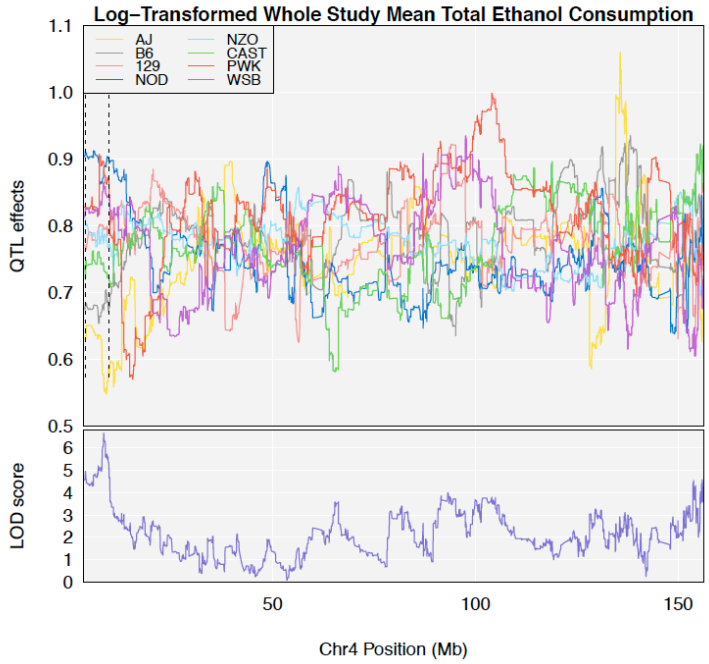
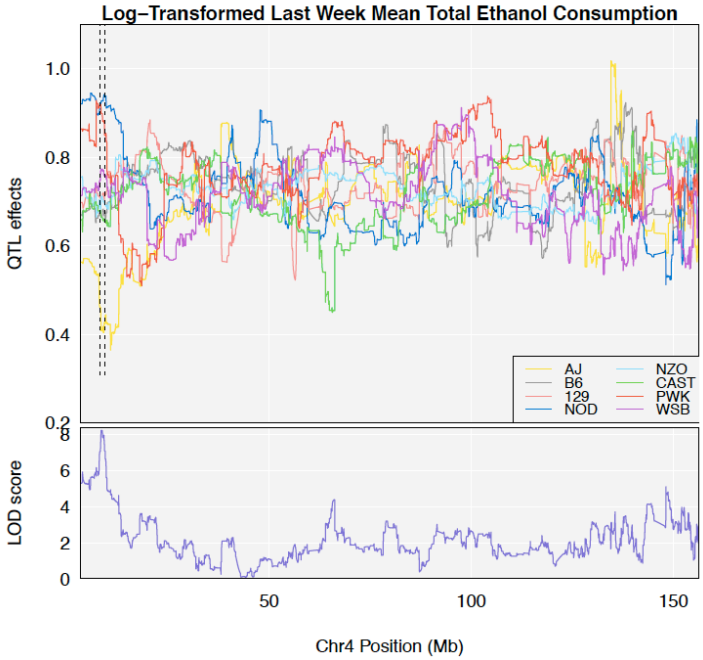
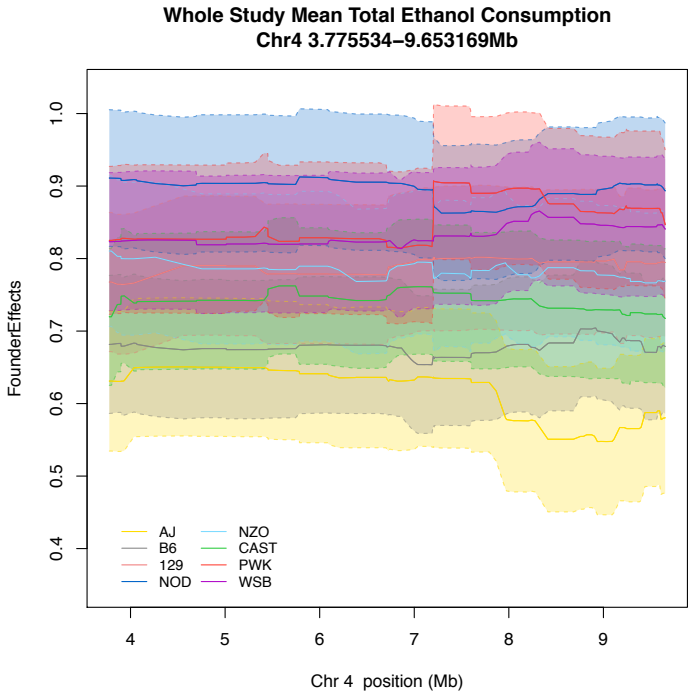
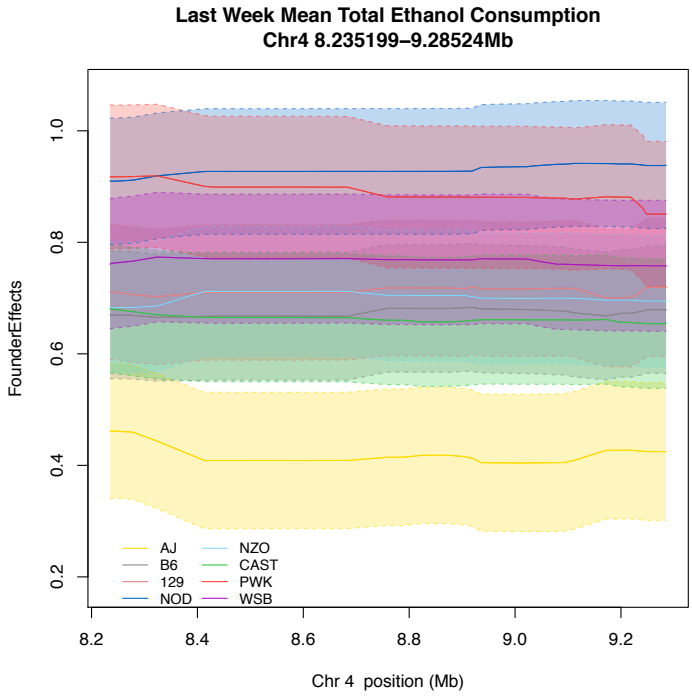


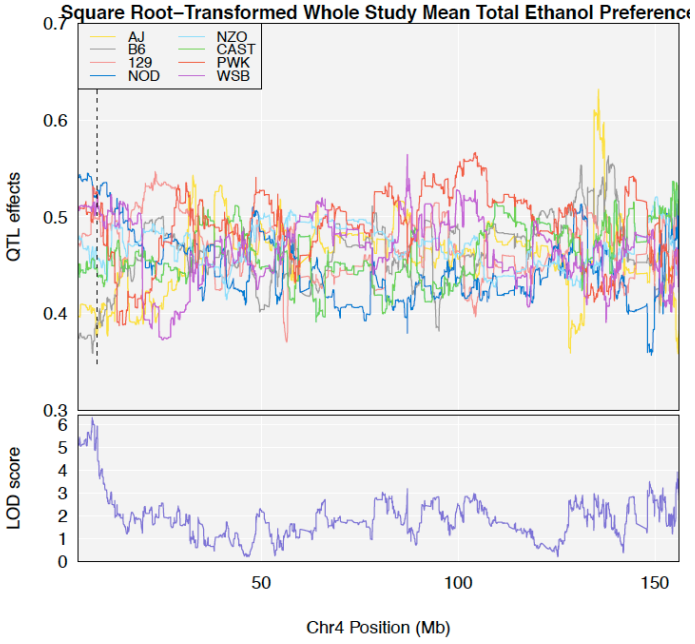
Figure 21: LOD plots for all ethanol drinking phenotypes (denoted in top right corner of each plot), across all chromosomes. Red and blue dotted lines denote empirical significance (corresponding to $p<0.05$) and suggestive thresholds (corresponding to $p<0.63$), respectively.

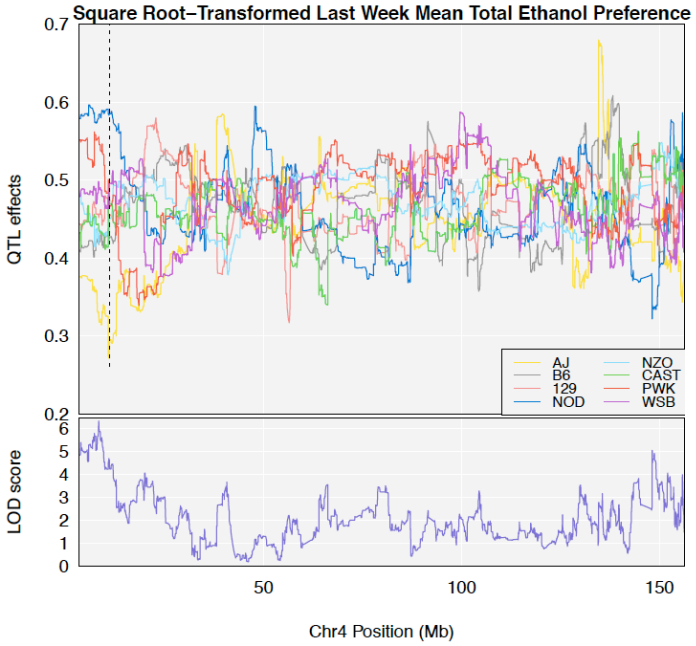


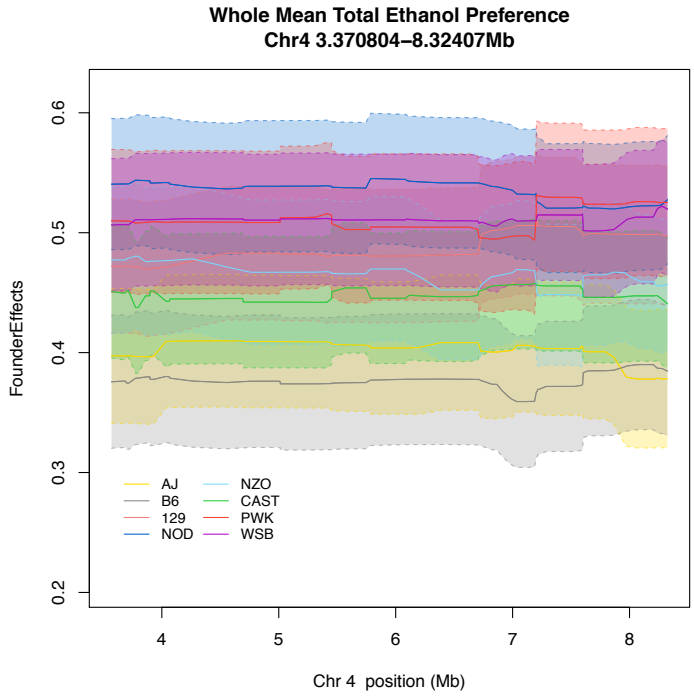


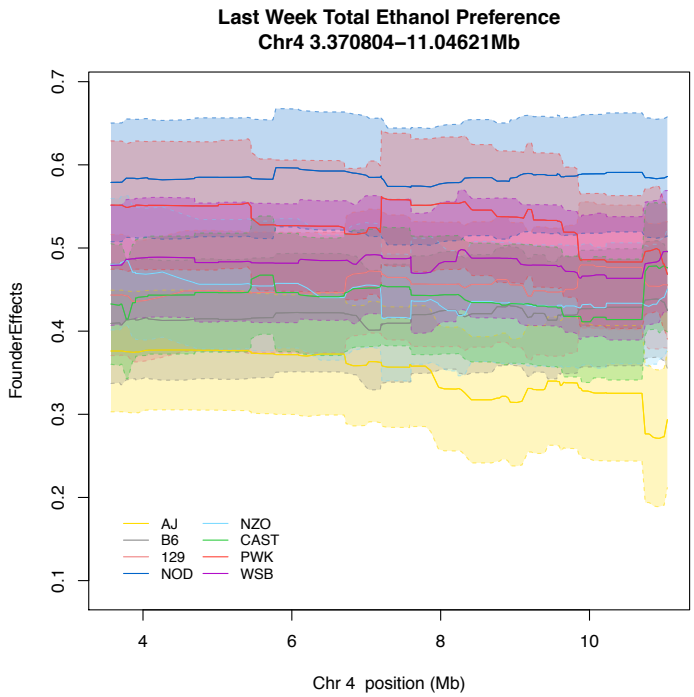


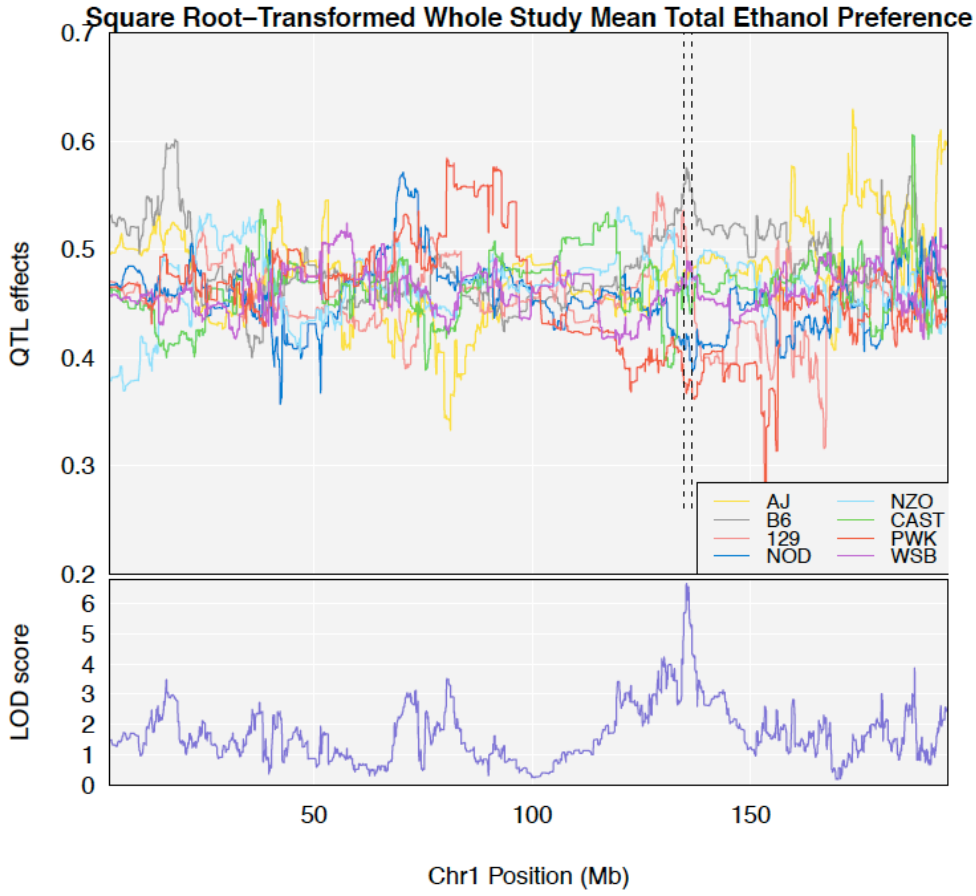


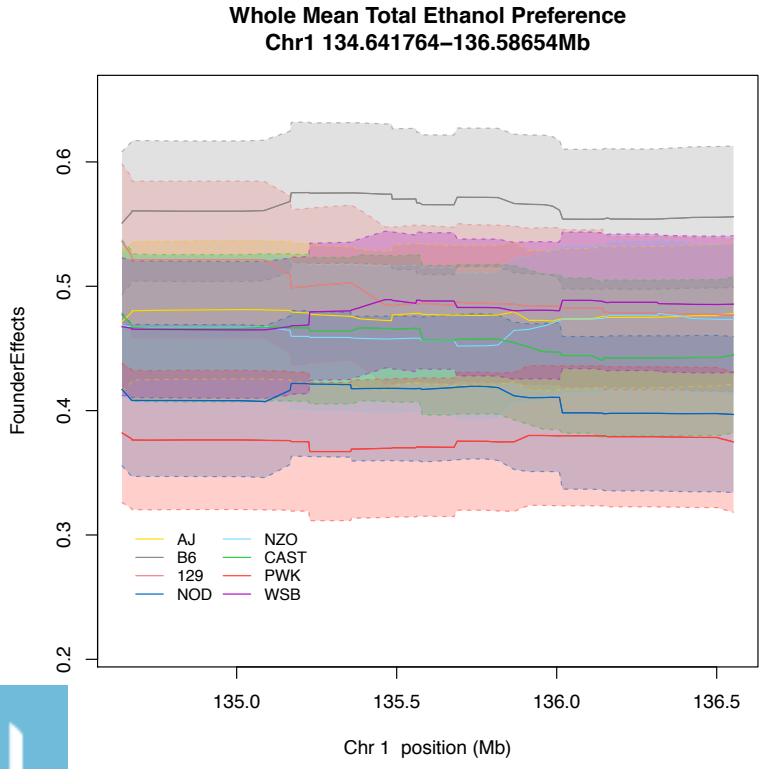


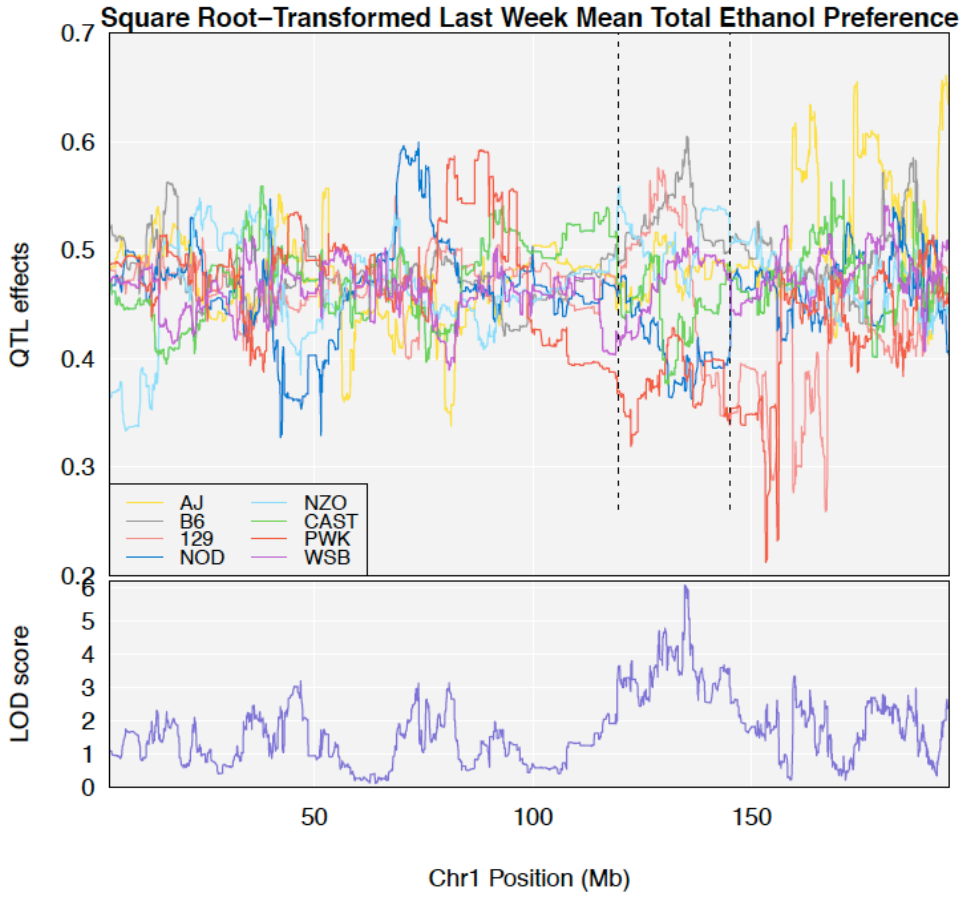


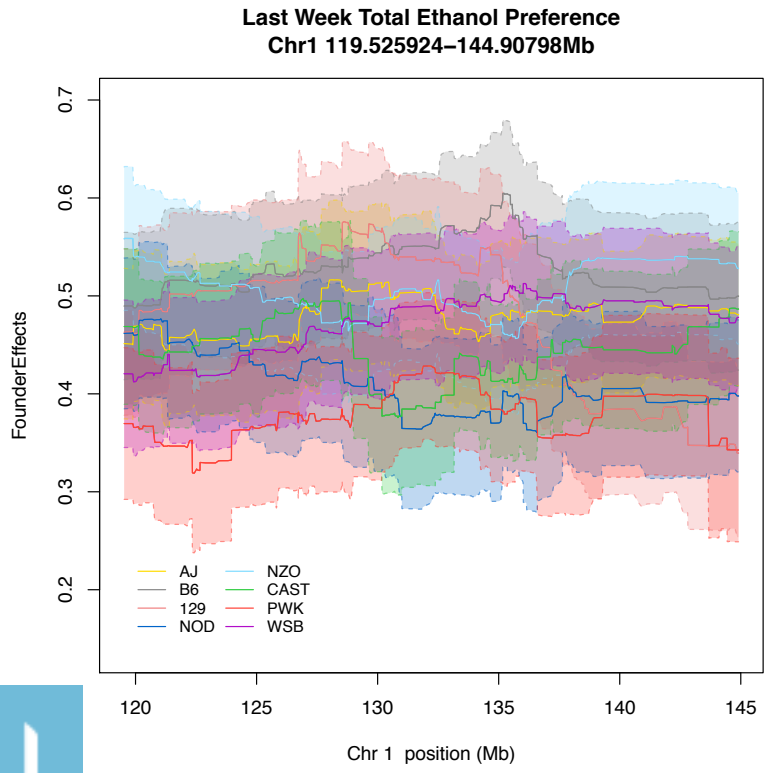


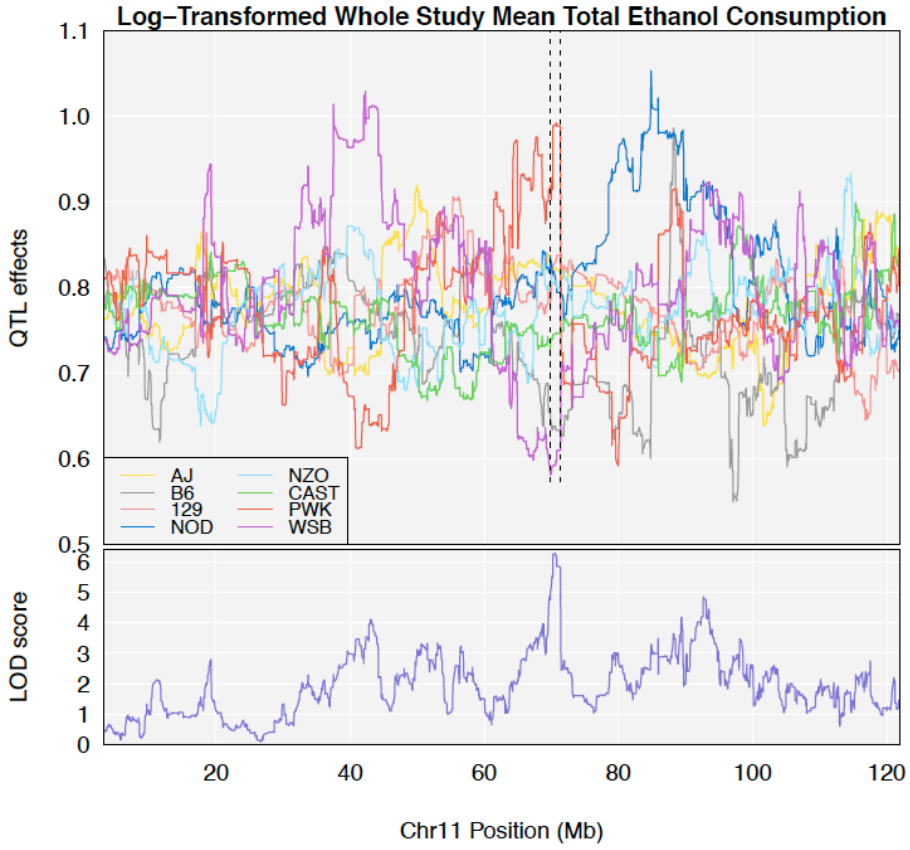


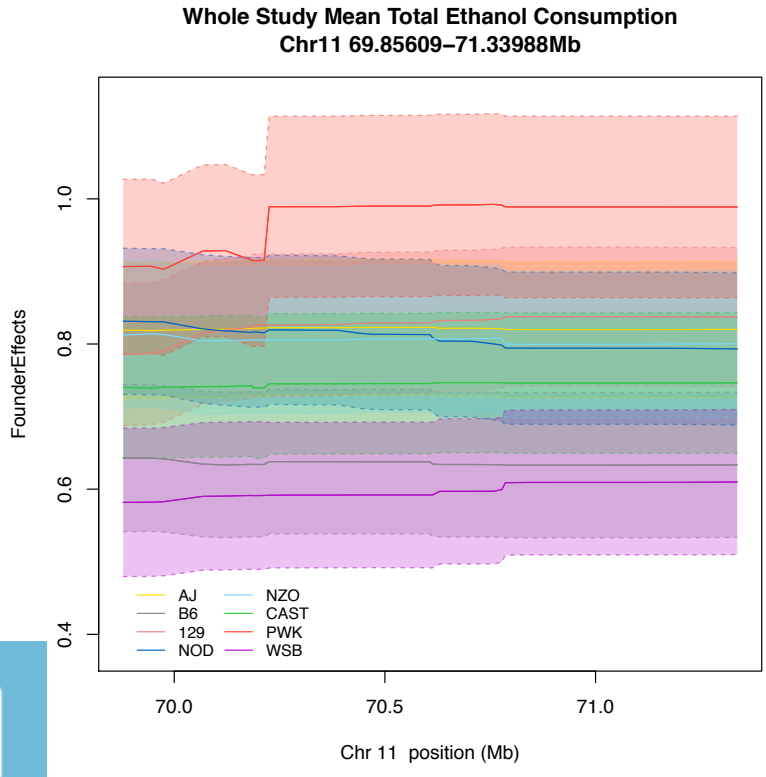


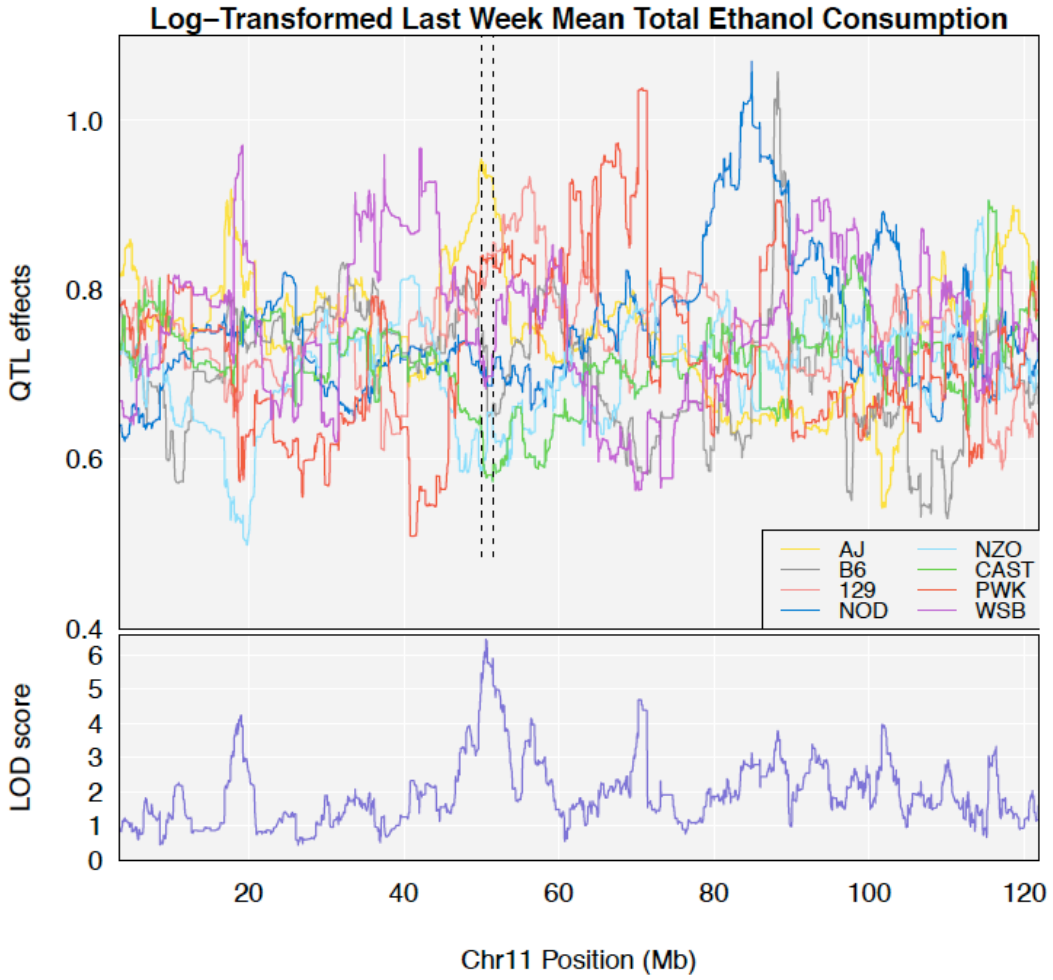


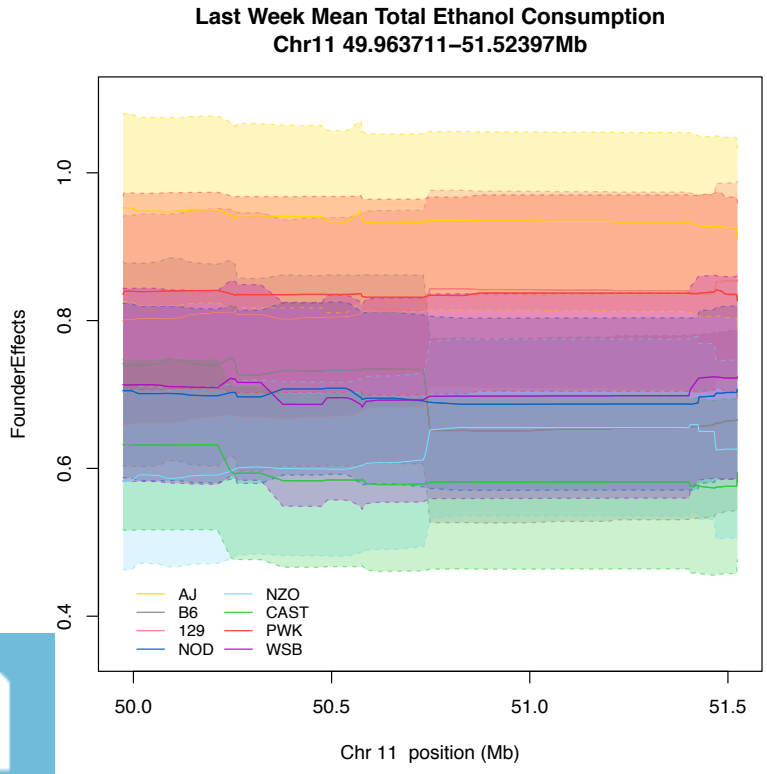


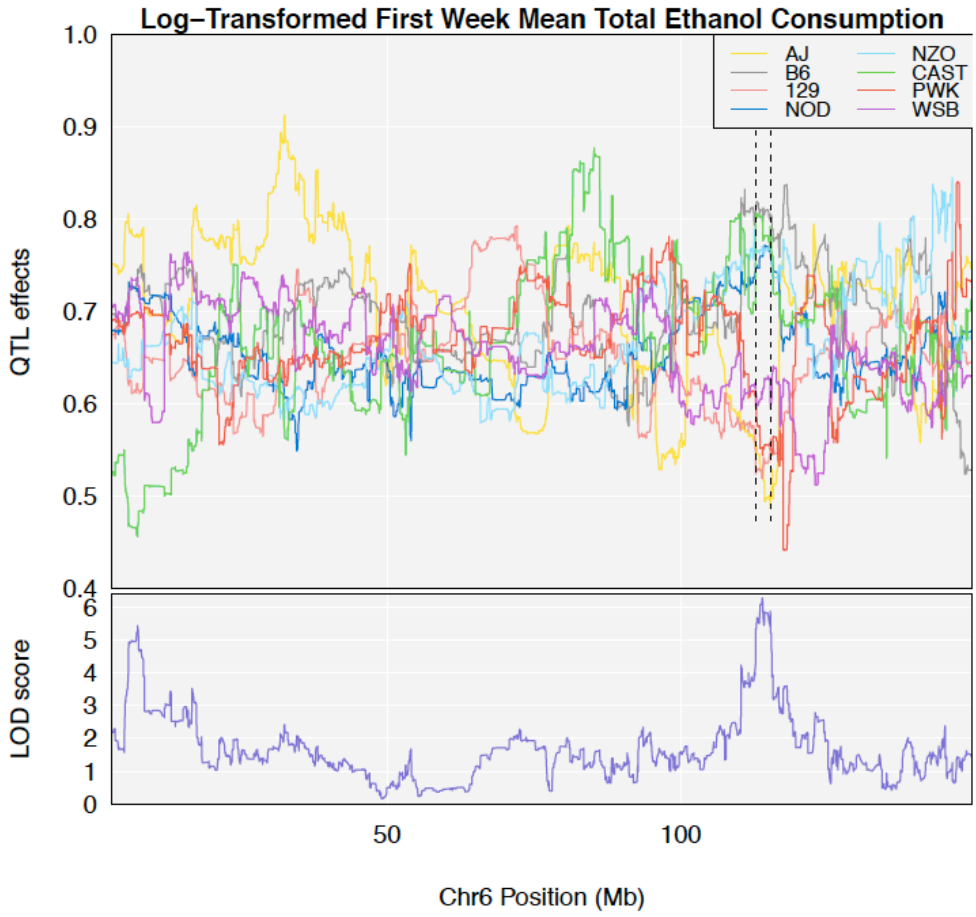


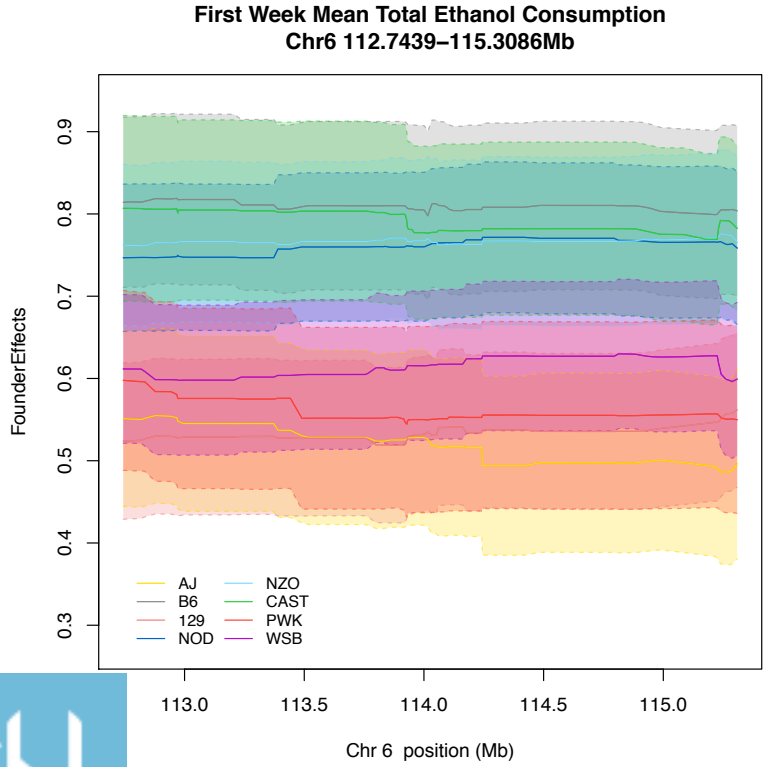


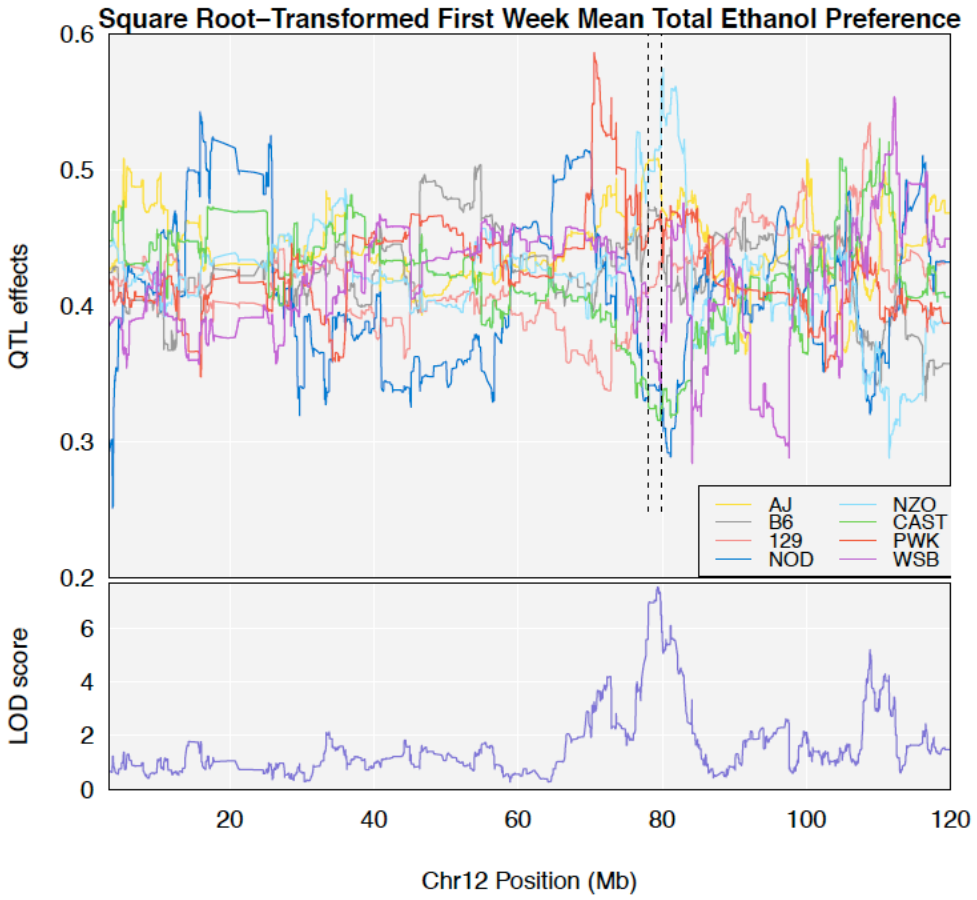


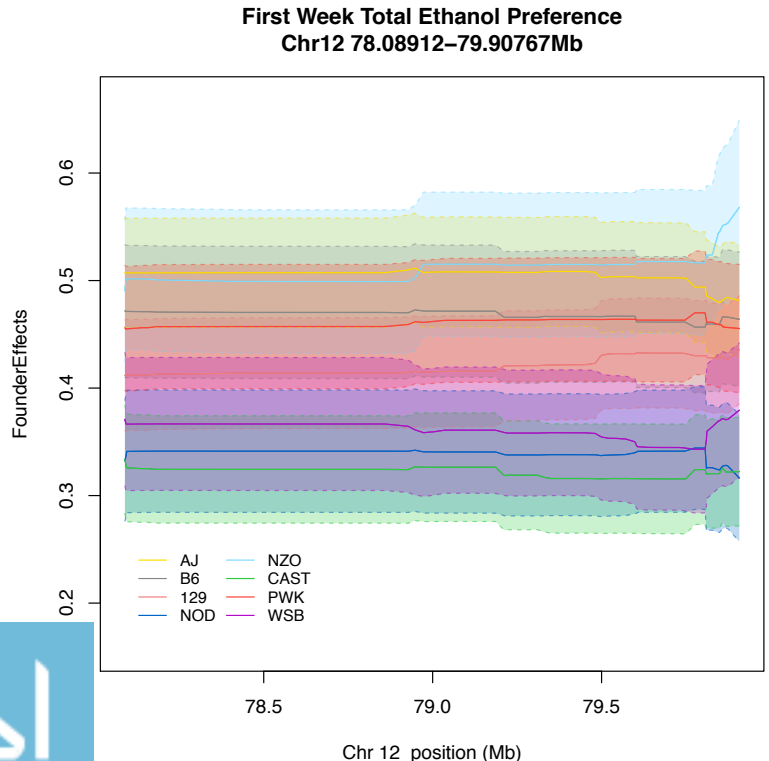


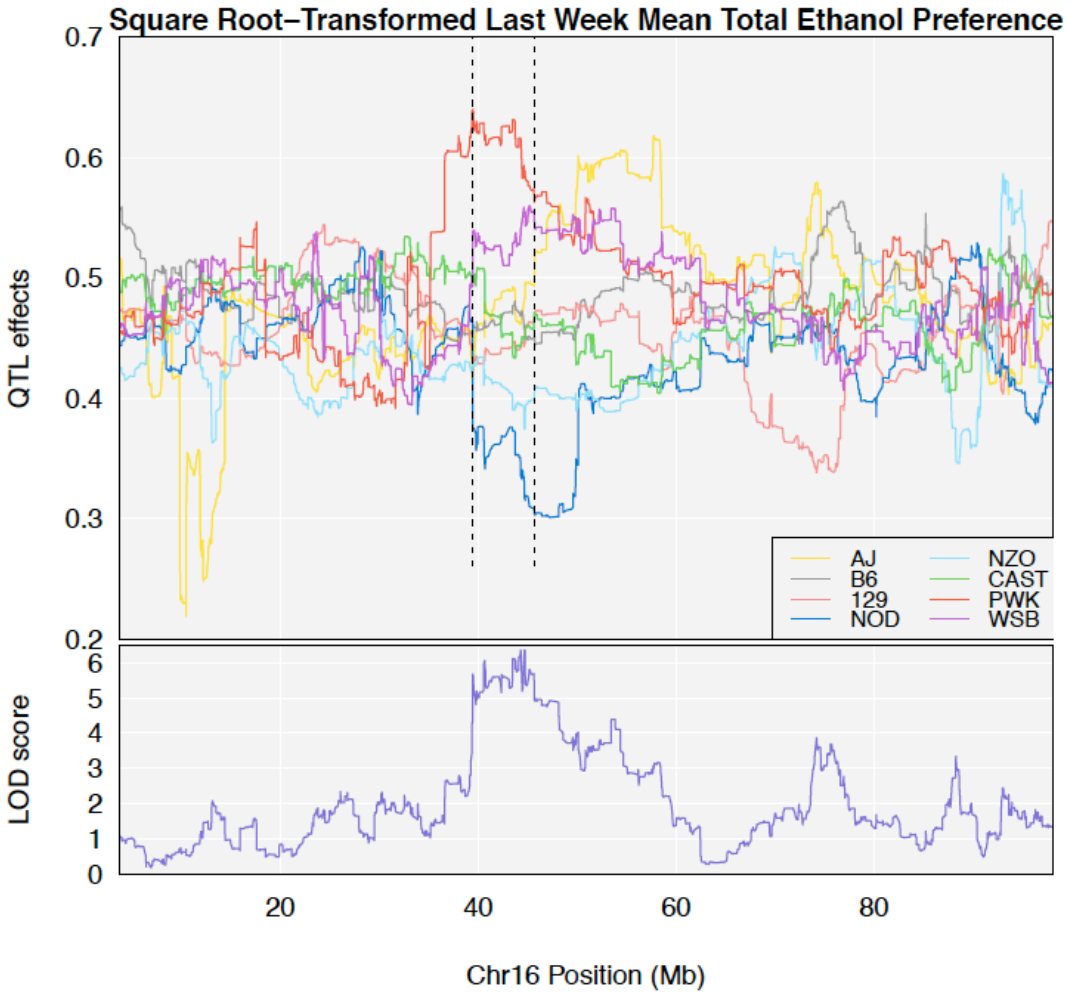


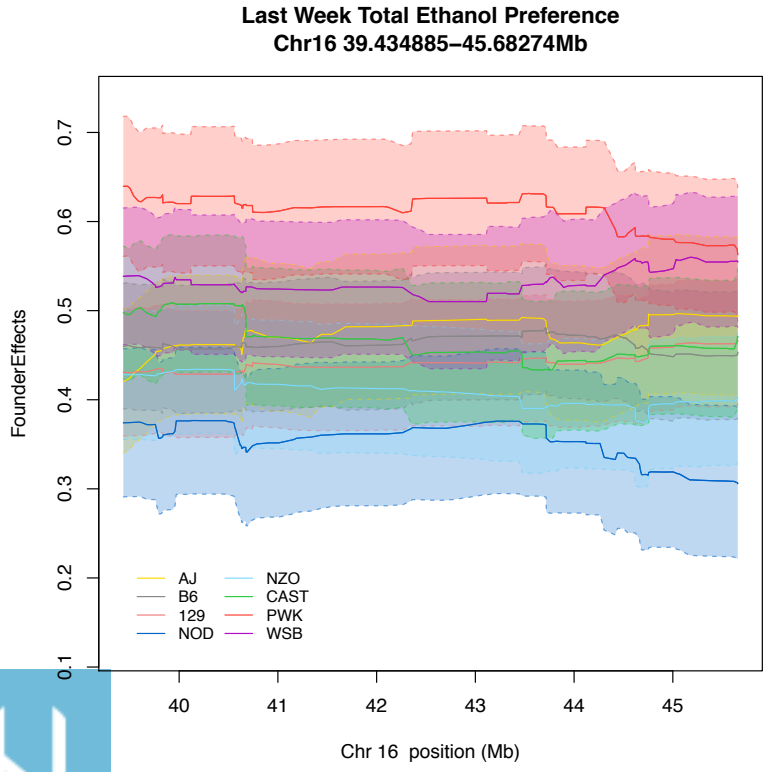


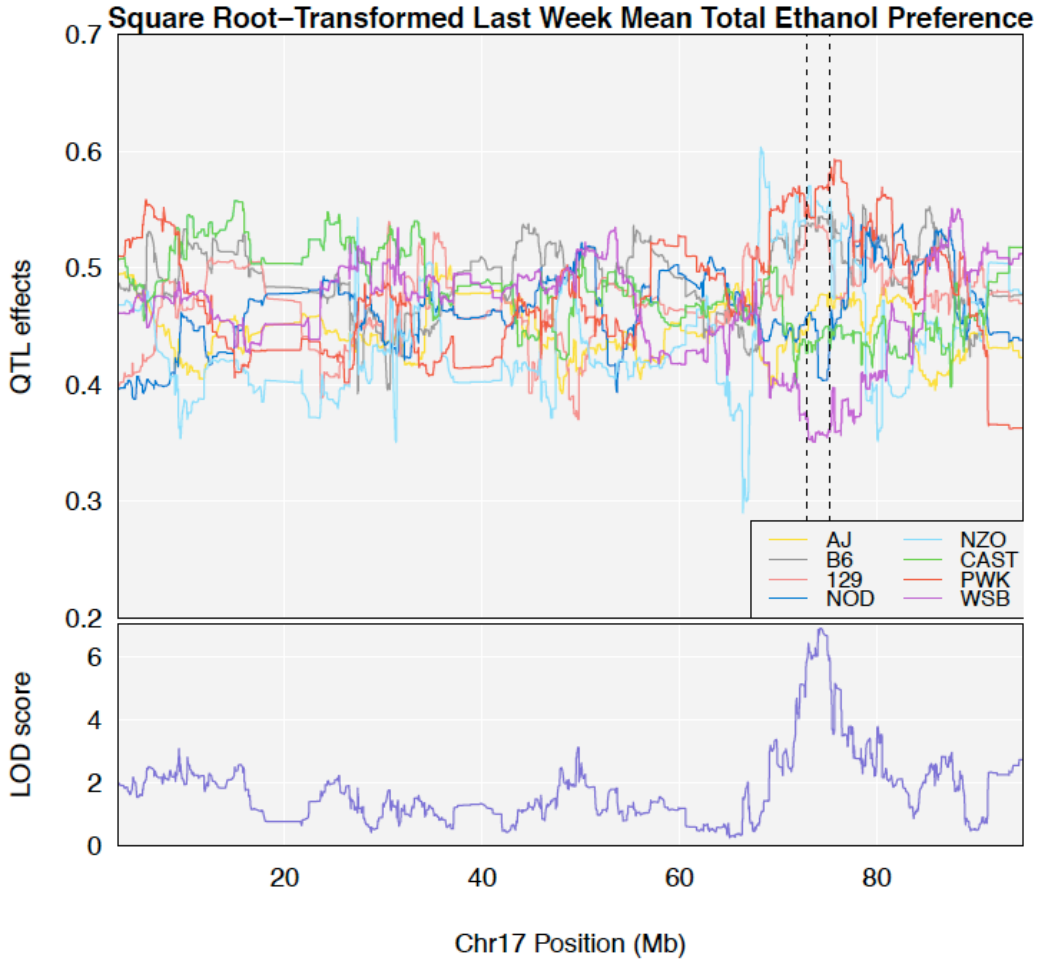


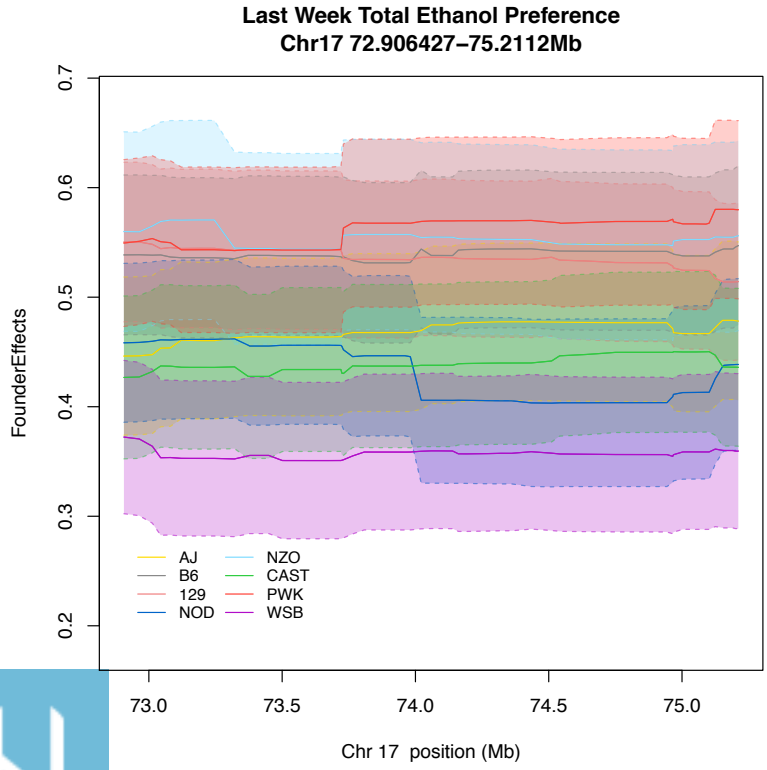


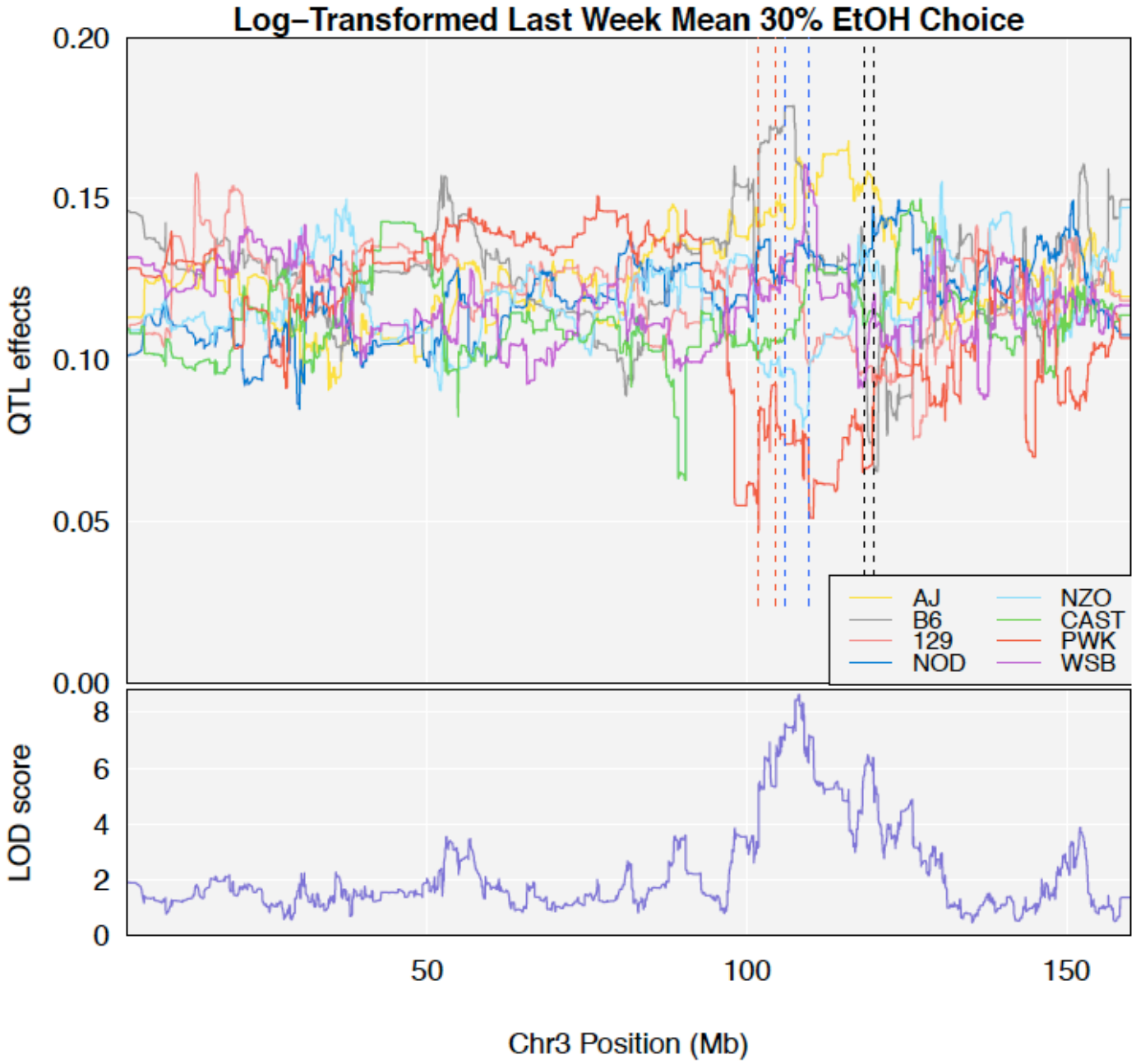


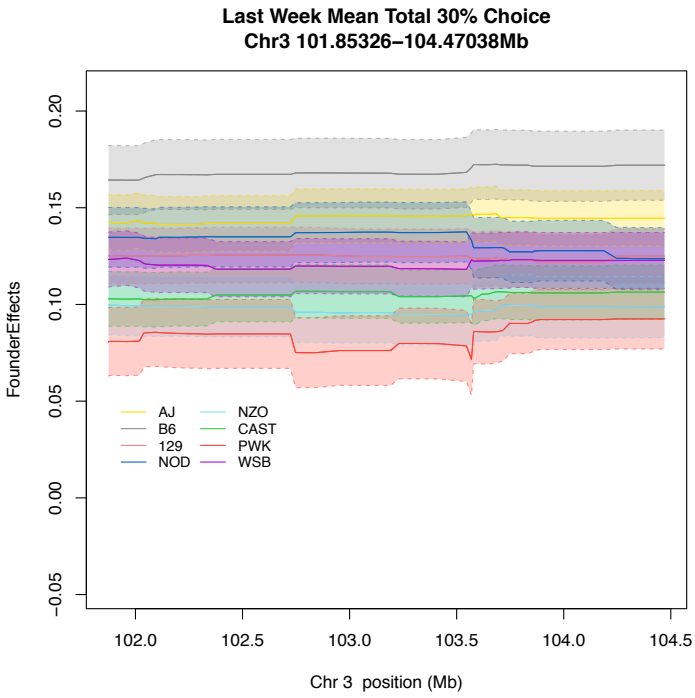


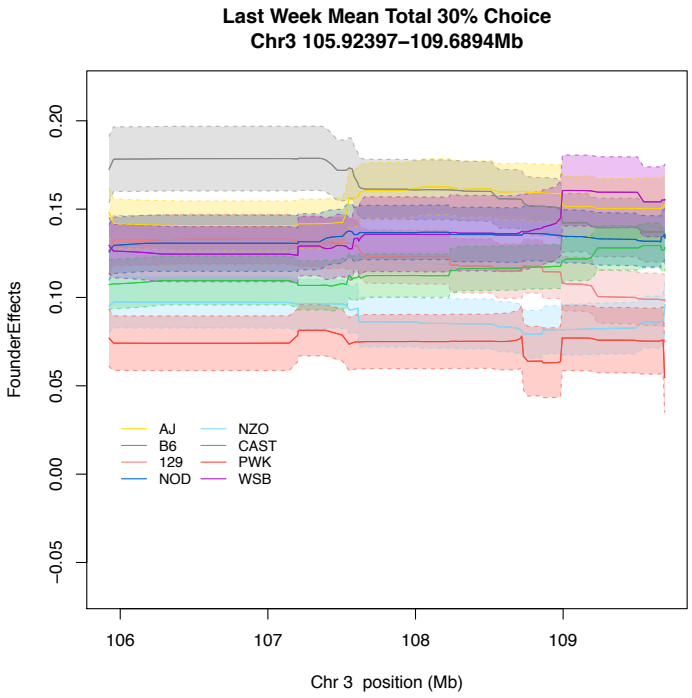


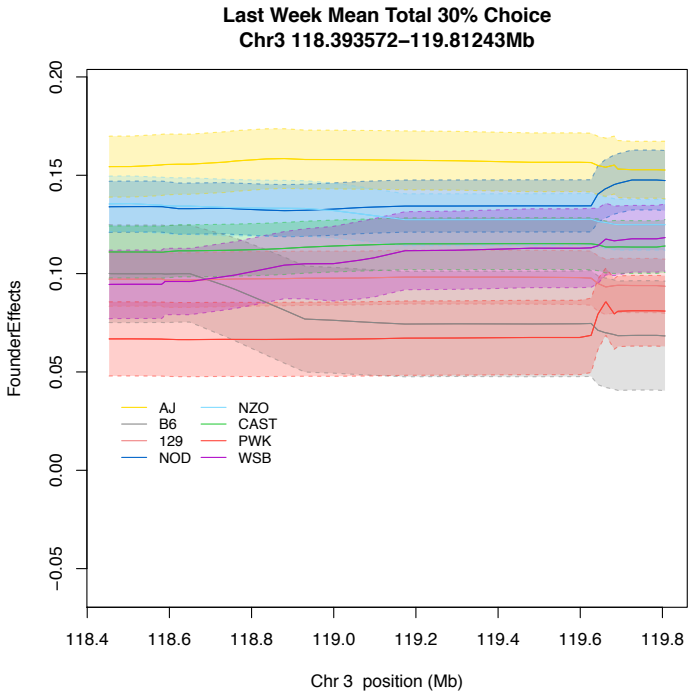


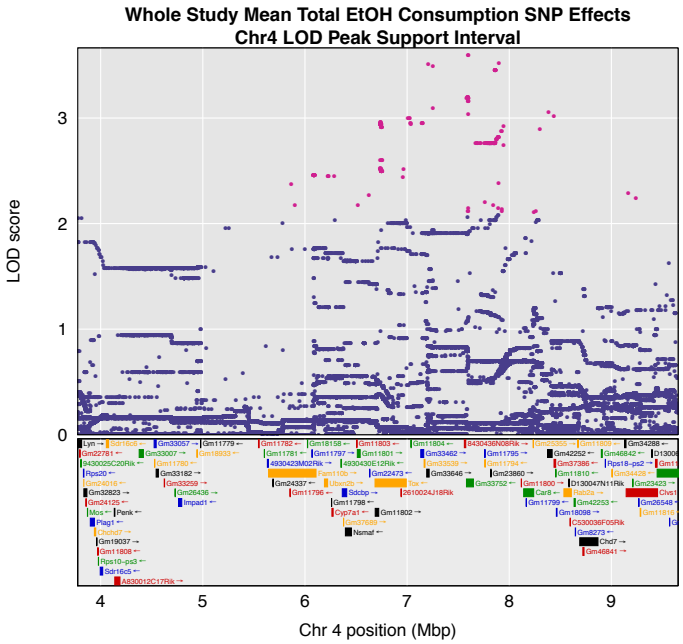


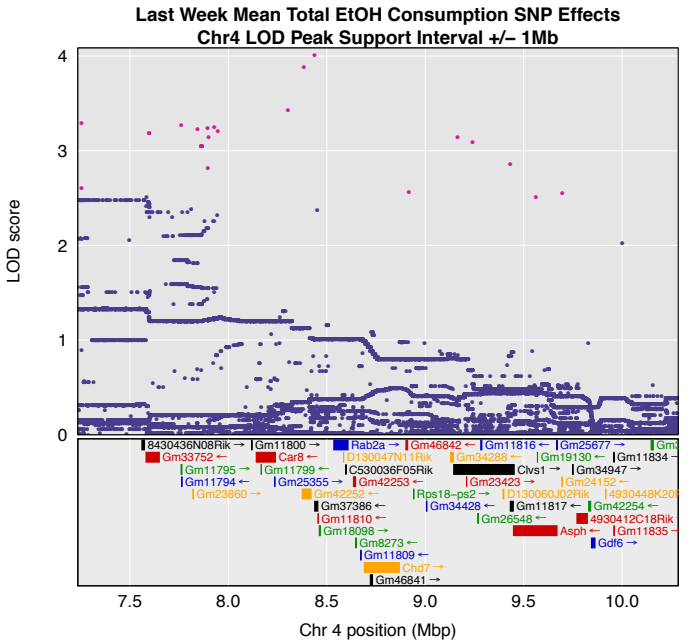


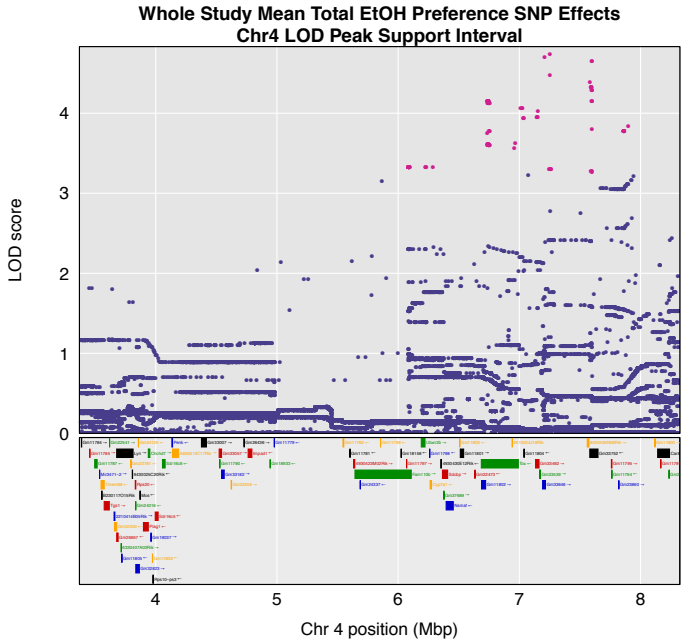


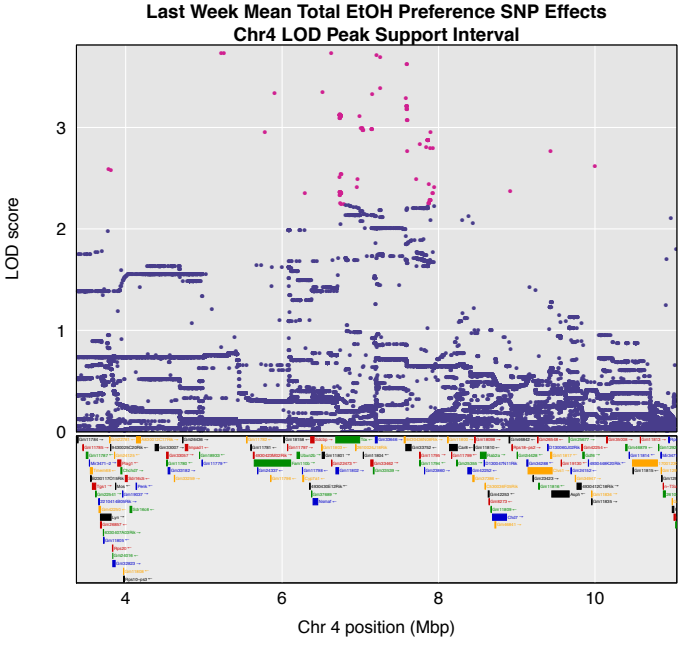


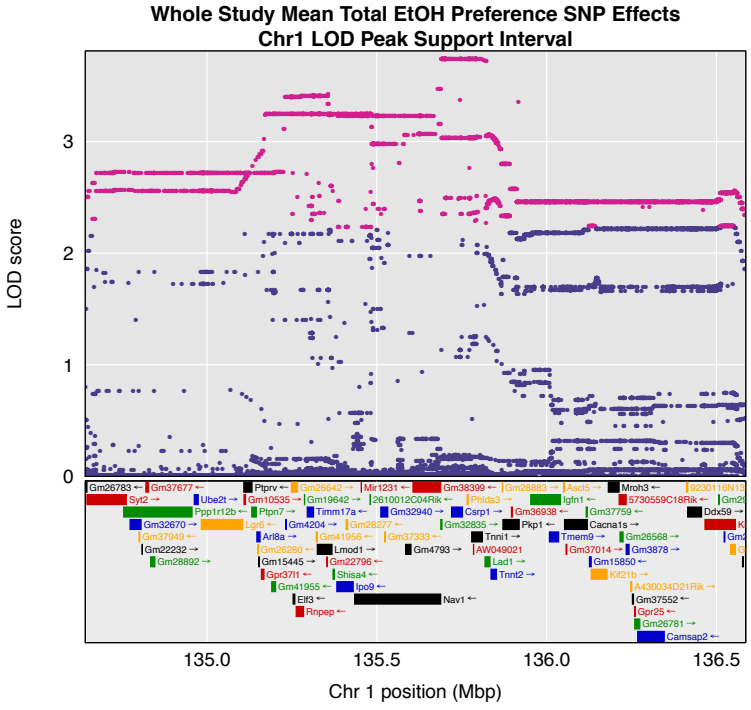


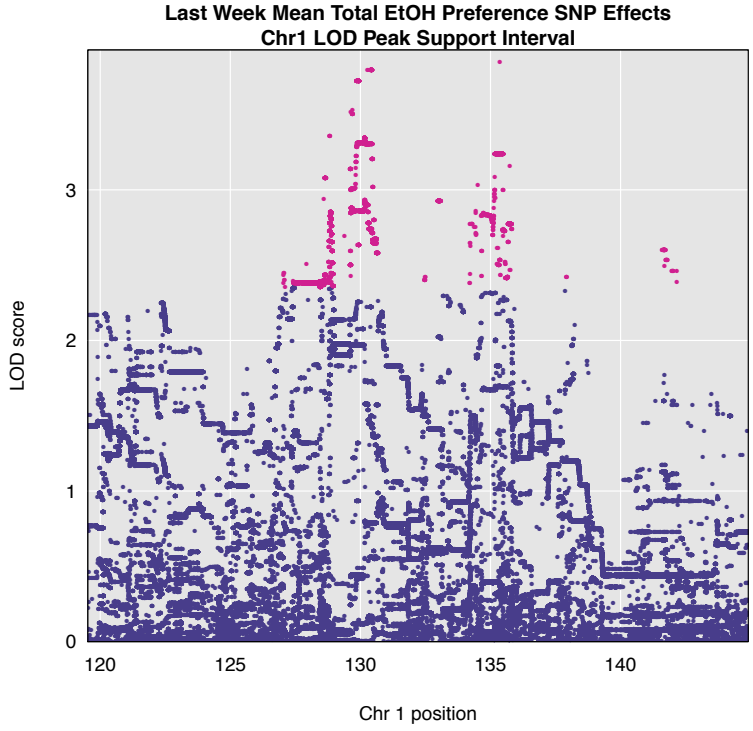


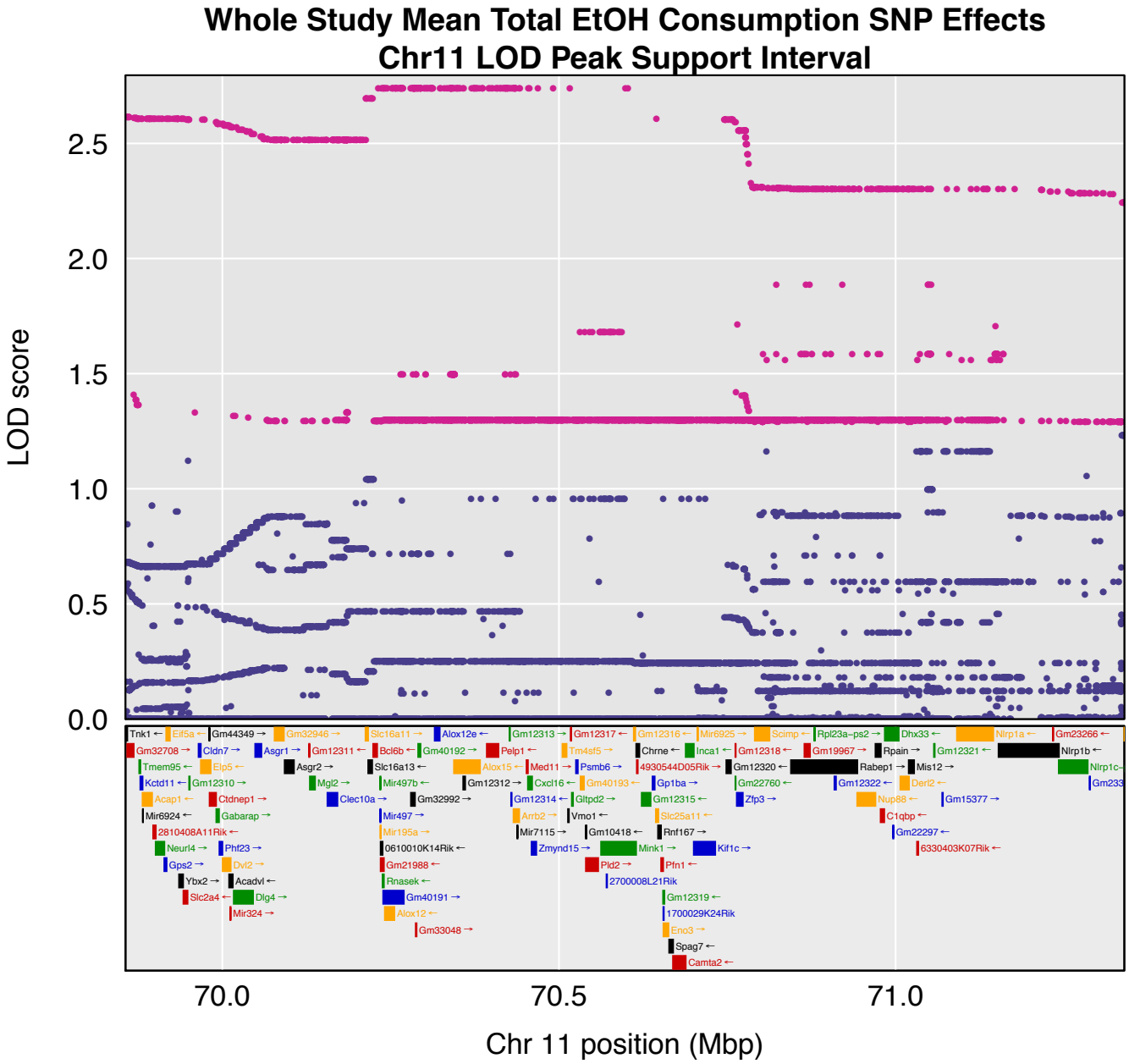


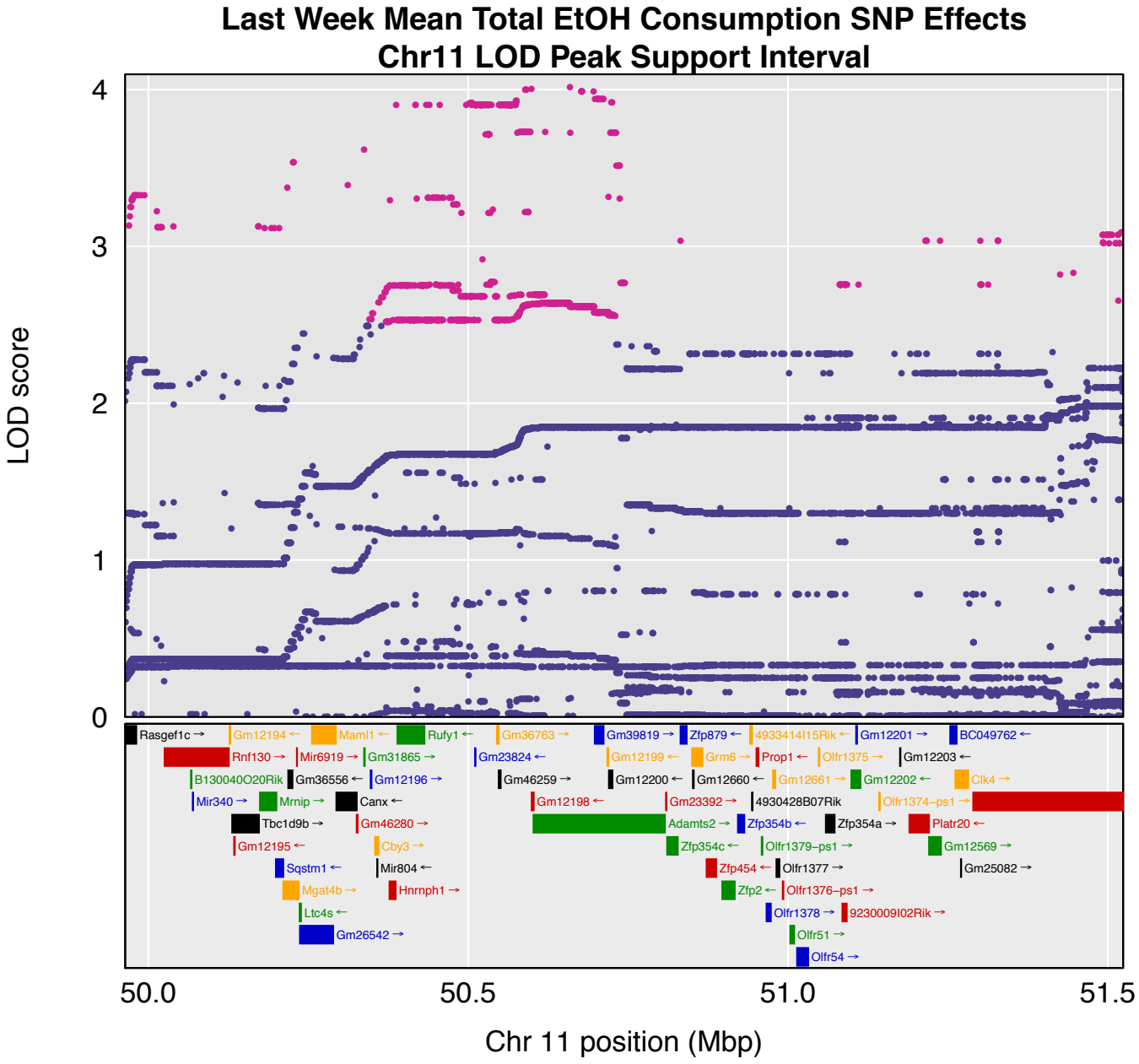


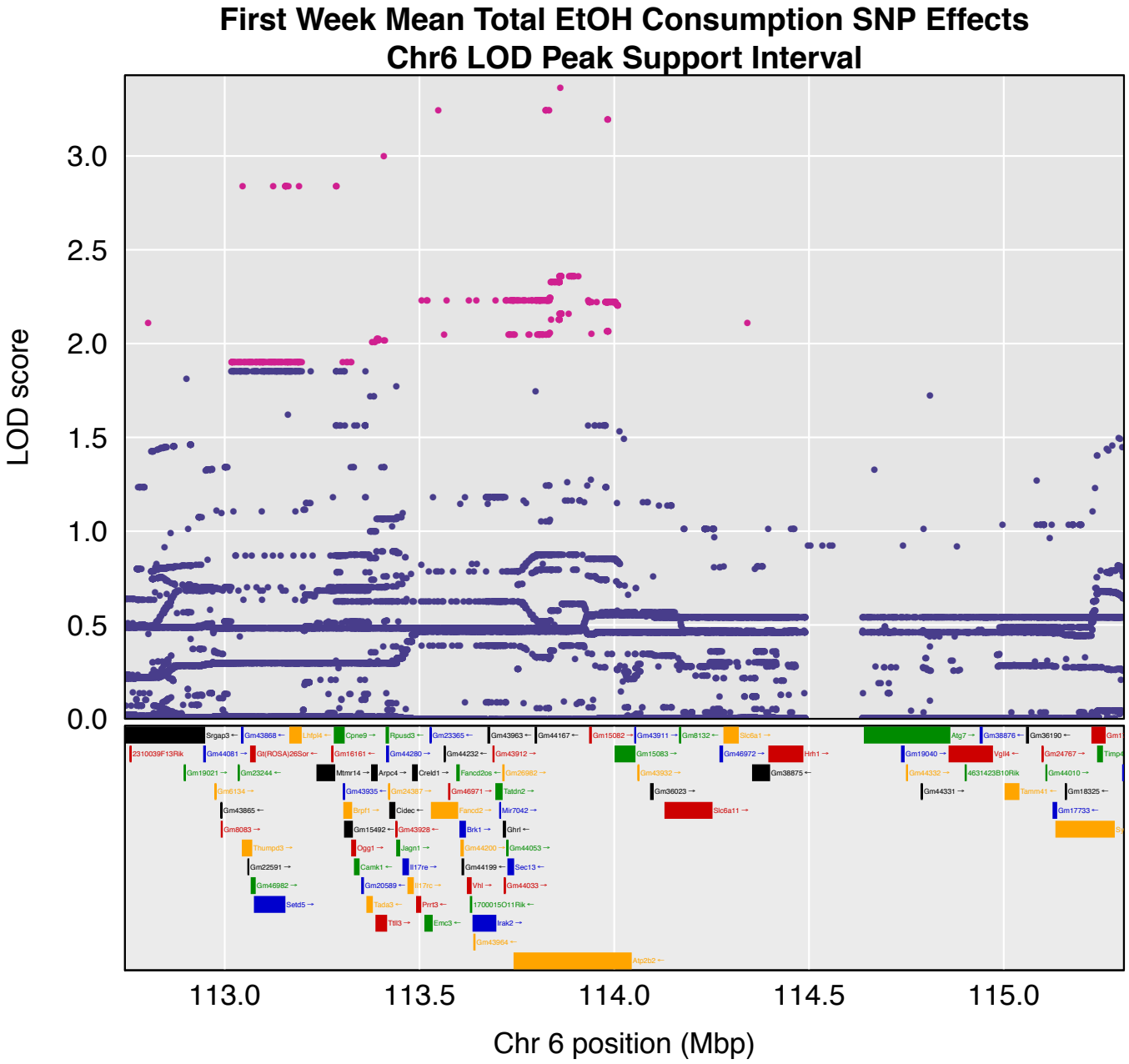


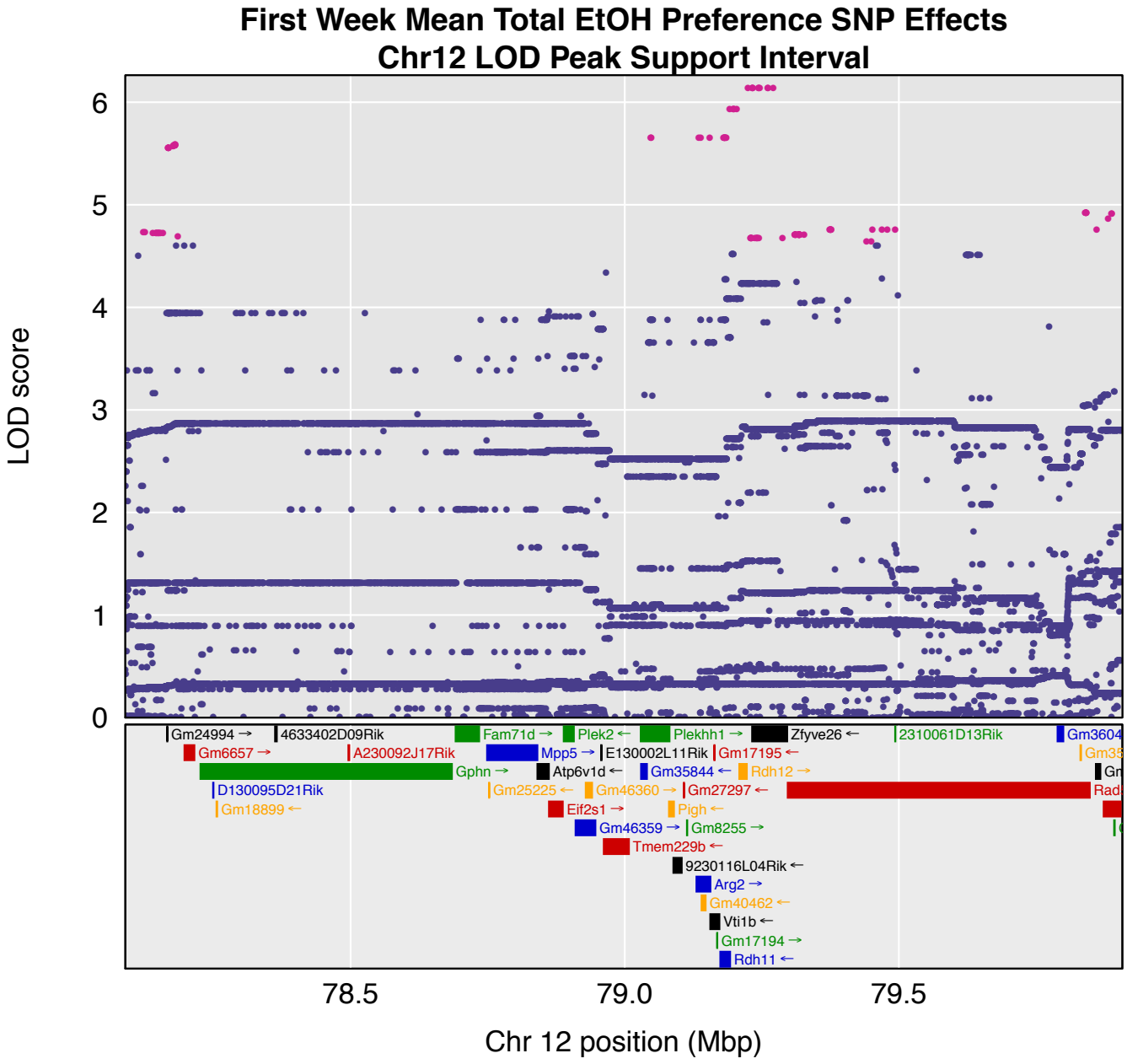


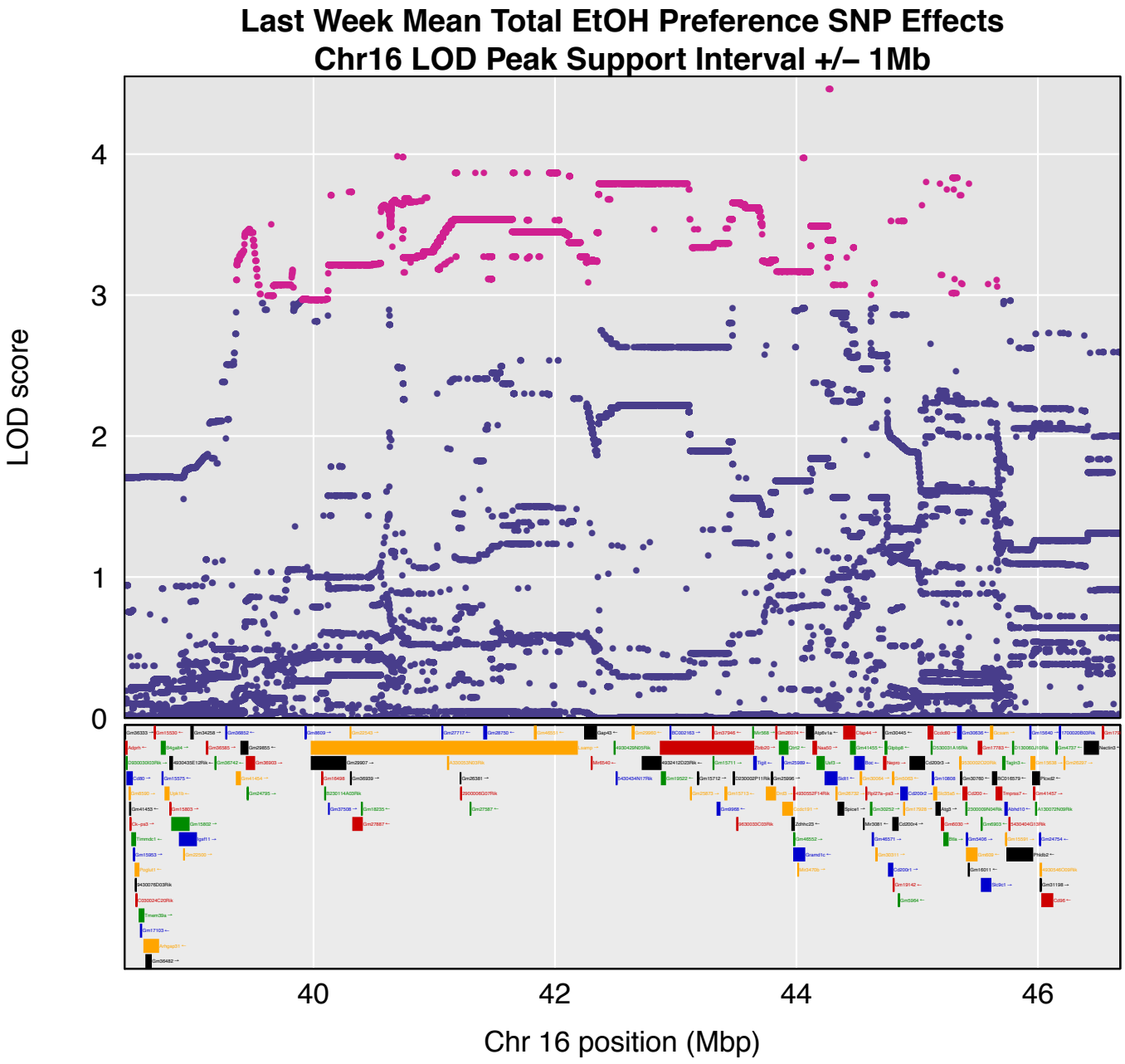


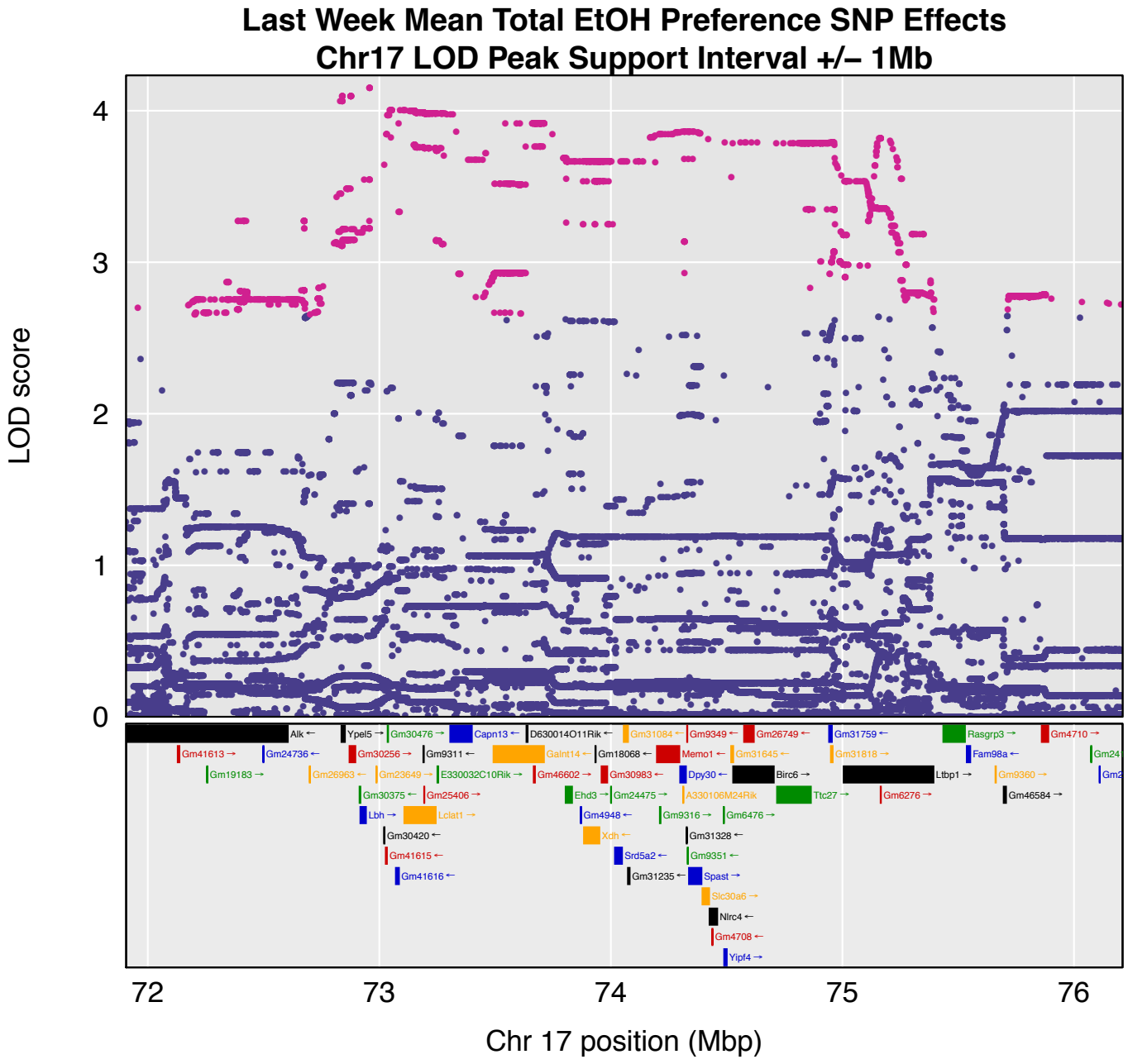


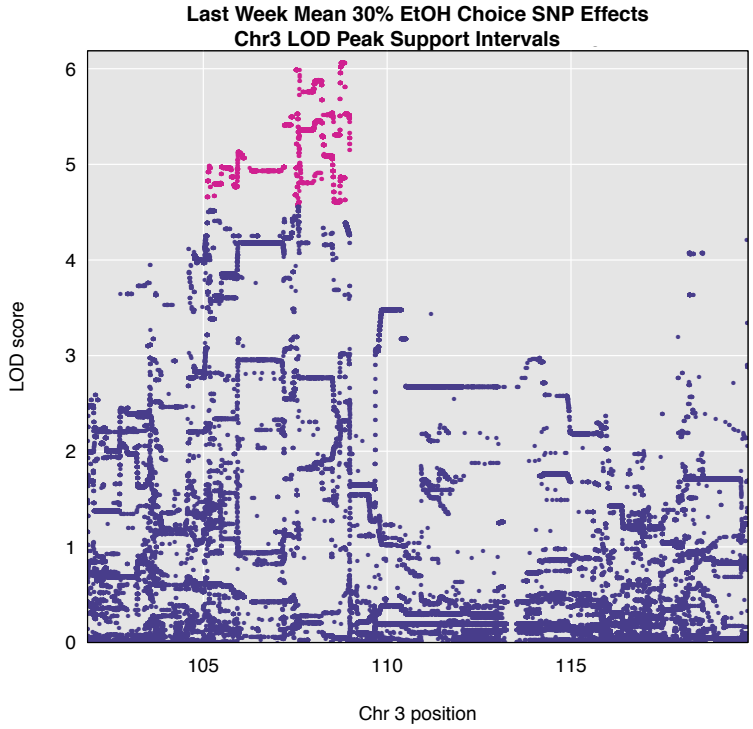


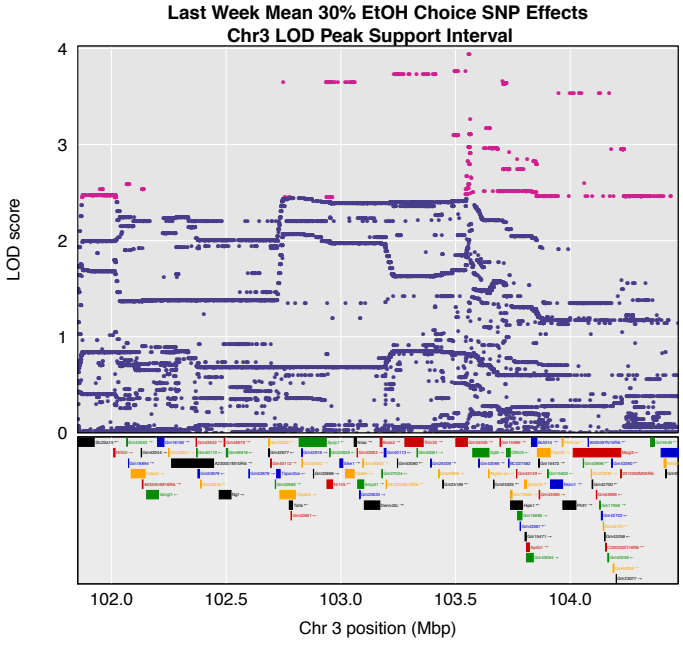


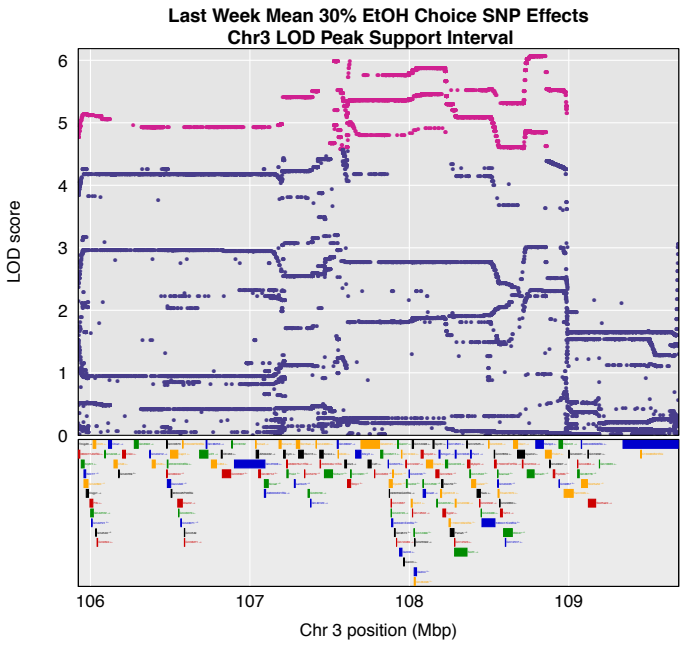












Discussion

Chromosome 4 Locus

To our knowledge the locus we identified for *Whole Study Mean Total Ethanol Preference* and *Consumption*, and *Last Week Mean Total Ethanol Preference* and *Consumption* on Chr4 has not previously been found to be associated with alcohol-related phenotypes, and is therefore a novel finding. Upon further investigation, three of the top candidate genes for the Chr4 locus were particularly interesting: *Car8*, *Tox*, and *Impad1*. *Car8* was the only top candidate gene present in this interval for all four of its related phenotypes. Its expression has been found to be associated with cerebellar ataxia, tremor, and dendritic growth and development in Purkinje cells (173-175). These functions could be relevant to the neuroplastic changes that occur in response to long-term ethanol abuse. Similar to *Car8*, *Tox* is evidenced to act as a transcription factor, regulating neurite outgrowth and neural stem cell proliferation during corticogenesis (or the development of the cerebral cortex) (176). It has also shown association with response to antipsychotics in human GWASs (170). Finally, human GWASs also indicate that *Impad1*, a gene involved in the hydrolysis of PAP (phosphoadenosine phosphate) into AMP (adenosine monophosphate) (177), is associated with not only alcohol dependence, but also response to SSRIs in depression treatment and hepatocyte growth factor levels (170). Given that depression and alcohol dependence are highly comorbid, and the metabolism of psychotropic drugs (such as alcohol or SSRIs) in the liver is important to response levels (as with ADH and ALDH enzymes), this gene is particularly intriguing.

Chromosome 1 Locus

The support interval of the locus we identified on Chr1, for *Last Week* and *Whole Study Mean Total Ethanol Preference*, overlaps with previously identified loci associated with ethanol preference and hypnotic sensitivity to high doses of ethanol in F2 and recombinant inbred mice, and with ethanol consumption in the GSCAN GWAS (86, 178). Five genes that fell within our *Last Week* support interval *SLC45A3*, *NUCKS1*, *RAB29*, *SLC41A1*, and *PM20D1* all contained variants that were considered to be in linkage disequilibrium ($r^2 > 0.3$) with the significant SNP rs823114 ($p = 2.31*10^{-9}$). *NUCKS1* also displayed significance in the gene-wise analysis of this study (Liu et al., 2019). However, our support interval was 48Mbp smaller (50Mbp vs. 2Mbp) than the (highly overlapping) previously identified loci. All five of the top SNPs for this locus are intron variants in the gene *Csrp1* (cysteine and glycine-rich protein 1), which may be involved in the development of muscle tissue (179). However, unlike other genes in this region, *Csrp1* does not appear to have been previously associated with psychiatric- or addiction-related traits. It is possible that intron variants in this gene could regulate the expression of genes up- or down-stream of it. Based on evidential support examined from the other studies mentioned above (Rhesus macaque chronic drinking anterior cingulate cortex and subgenual cortex gene expression data, acute-ethanol-treated BXD brain-region-specific gene expression data, DO liver eQTL, and human alcohol-related phenotype GWAS results), this locus contained several top candidate genes: *Lmod1*, *Nav1*, *Pm20d1*, *Slc41a1* and *Nucks1*. *Lmod1* and *Nav1* were captured in both the *Last Week* and *Whole Study* QTL. *Lmod1*'s encoded protein, Leiomodin-1, is an actin-binding protein that is involved in the nucleation of actin filaments (180), and

Nav1's protein, neuron navigator 1, is also associated with microtubule-mediated neuronal migration (181). Actin is often thought to be involved solely in muscle function, but it also serves as scaffolding for microtubules as they migrate to the to the end of the axon to facilitate cone outgrowth (182). Although the *Whole Study* QTL support interval does not encompass the final three top candidate genes for this locus, it is still tagging the same chromosomal region. Therefore, for thoroughness, all genes within the larger support interval should be considered, as done with the Chr4 locus. *Pm20d1*, *Slc41a1* and *Nucks1* all contain significant SNPs for *Average Drinks per Week* in the GSCAN GWAS (61). The association of three genes within one region in each of two studies is not a coincidence. This region is, in fact, conserved across the mouse and human genomes. This suggests that our analysis, which is investigating the mouse behavioral model for average drinks per week (after long-term alcohol exposure), replicated a locus identified in a human GWAS.

Nucks1 is a particularly interesting gene in this region, as it has shown association with Alcohol Consumption and Parkinson's Disease in human GWAS (61, 170), and its expression levels (in brain or blood) are significantly associated with Parkinson's, Alzheimer's Disease, Bipolar Disorder, and depression, all psychiatric and neurological disorders (140, 183-185). The exact function of *Nucks1* is unknown. However, there is evidence of its involvement in homologous recombination and inflammatory immune reactions. It is expressed ubiquitously throughout the body, and possesses phosphorylation sites for cyclin-dependent kinases and casein kinase II, both of which play regulatory roles in the cell cycle (186).

Chromosome 3 Loci

QTL previously identified for ethanol-conditioned taste aversion and body temperature change in response to methamphetamine, overlap all three of the loci identified in the present study for Chr3 for *Last Week Mean Total 30% Ethanol Choice* (68, 71). The most proximal two loci were directly juxtaposed, with support intervals separated by only 1.45Mbp, and were positioned over the same LOD score peak. Therefore, while looking for candidate genes, we considered these two loci to be part of a single QTL, and considered all genes within and between (i.e. within the 1.45Mbp gap) them. Even when considering both of these loci together, their combined support interval is 42.2Mbp smaller (7.8Mbp vs. 50Mbp) than those identified in the aforementioned previous studies. None of the analyzed variants in the more distal of these two intervals followed the expected haplotype effects. However, it could potentially contain a gene affected by a variant in the most proximal interval. It could also possess other variants that may affect the same gene, but with a smaller effect size than the variant located in the more proximal locus. The top-scoring SNP for this region (rs49087152) is classified as a nonmediated decay transcript for the gene *Aknad1* (AKNA domain containing 1), about which very little is known, although one study suggests that it is regulated by TNF-α, which increases in expression during muscle degeneration (187). The SNPs with the next two highest scores (rs217960262 and rs244300971) are downstream variants of this gene. SNPs rs246339091 and rs213764891 are a non-coding transcript variant and intron variant of the gene *Stxbp3* (syntaxin binding protein 3), respectively. Interestingly, *Stxbp3* encodes a syntaxin protein. This type of protein binds with synaptotagmin, a type of protein that is encoded by one of the top candidate genes for this region, *Syt6*. This gene translates into the synaptotagmin 6 protein, which binds to calcium

and participates in anchoring vesicles to the presynaptic membrane in the process of exocytosis of neurotransmitters, and plays a role in neurite outgrowth. Given its highly relevant function, it is not surprising that is has displayed association with elicit drug use and Bipolar Disorder and Schizophrenia in humans (170).

Three other top candidate genes stood out after further investigation: *Gnai3*, *Tspan2*, and *Sort1*. *Gnai3* encodes the alpha subunit of G proteins, which are involved in intracellular signaling cascades, and it has been found to be associated with unipolar depression in human GWAS (170). *Tspan2*'s encoded protein, tetraspanin-2, is also involved in signal transduction, specifically for signals involved in cell development and motility (https://www.ncbi.nlm.nih.gov/gene/10100#reference-sequences). Not only does it display a significant association with the neurological disorders Schizophrenia and migraines in human GWAS (170), but its expression has also been shown to be associated with impulsivity and neuroinflammation in mice (188, 189). High impulsivity is a hallmark symptom of both inebriation and long-term alcoholism, and neuroinflammation is known to be associated with chronic alcoholism (8) (Kohno et al., 2019; Tyler et al., 2019).

Sortilin, encoded by *Sort1*, is a cell-surface and Golgi apparatus protein. It has several functions including: being responsible for the endocytosis of neurotensin; sorting proteins (including neurotensin) for transport from the Golgi to lysosomes and endosomes; and moderating apoptosis of neurons via its involvement in BDNF and NGFB endocytosis (190) (191) (192). Human GWAS shows association between this gene and intelligence (170). *Sort1*-deficient mice exhibit increased anxiety-like behavior (193), however, increased levels has been shown to be associated with depression in humans (194). It also plays a role in

NGF-moderated neurodegeneration, autoimmune-related neuroinflammation, familial essential tremor, Alzheimer's, and frontotemporal dementia (195-197) (198) (199).

Finally, the most distal of the three bQTL on Chr3 for *Last Week Mean 30% Ethanol Choice*, was located above a LOD peak that appeared to be distinct from the one spanned by the more proximal two loci. Therefore, this locus was analyzed as a distinct QTL. Its support interval was only 1.4Mbp wide, compared to the 50Mbp intervals previously identified. Although this locus contained variants with SNP-based LOD scores > 4, none of the tested variants in this region followed the expected haplotype patterns in the progenitor strains. Therefore this interval is to be interpreted with more skepticism than the other, more proximal loci. Nonetheless, both of the top candidate genes for this QTL (*Dpyd* and *Ptbp2*) have exhibited associations with a wide array of neurocognitive and psychological traits in previous studies. *Dpyd* encodes the protein dihydropyrimidine dehydrogenase, which is involved in the catabolism of pyrimidines (uracil and thymine nucleotides) (131). Despite its regulatory function, rare mutations in this gene have been found, and are related to schizophrenia and intellectual disability (200) (201). Furthermore, human GWAS have identified associations between normal variants in this gene and Schizophrenia, as well as several other relevant phenotypes, including: intelligence, nervousness, sleep duration, educational attainment, cognitive function, risk-taking behavior, social interaction, reaction time, and borderline personality disorder (2, 170). *Ptbp2* is also associated with Schizophrenia, social interaction, and reaction time, as well as headaches, in human GWAS. This gene's protein also has a pyrimidine-related function: binding to poly-pyrimidine strands of mRNA introns to regulate splicing, specifically in the brain (202, 203). This protein plays a role in neurogenesis and axonogenesis, and rare mutations in this gene are

associated with autism spectrum disorder and infantile neuroaxonal dystrophy (203, 204). Like *Car8*, *Tox*, *Zbtb20*, *Spast*, and *Syt6*, *Ptbp2* be specifically involved in the neuroplastic changes in the brain that are so heavily involved in the development and maintenance of alcohol dependence (15).

Because the loci identified in other studies had such large support intervals, it is unclear which of our loci is most likely tagging the same affective genes. However, the previously identified QTL were discovered in BXD samples. This means that a polymorphism between C57BL/6J and DBA/2J must have contributed to these effects. Of the top candidate genes for the two more proximal loci, *Syt6* (which falls within the middle interval) and *Dpyd* (which falls in the most distal interval) are the only ones that contain polymorphisms between C57BL/6J and DBA/2J (205). However, variants that fall outside the predefined realms of a gene (within a certain number of Mbp) can and do affect the transcription of genes farther down- or up-stream of them. Additionally, none of the analyzed variants in *Syt6* or *Dpyd* match the expected founder haplotype patterns for the middle or most distal interval, respectively. Therefore, although this provides further evidence that polymorphisms within these two genes are affecting *30% Ethanol Choice*, we cannot definitively rule out the other top candidate genes in these regions.

Chromosomes 16 and 17 Loci

The two loci that showed suggestive association with *Last Week Mean Total Ethanol Preference* are located on Chromosomes 16 and 17. Both of these loci overlap with QTL previously shown to be associated with methamphetamine-induced body temperature change sensitivity (just as the loci we identified on Chr3), and for locomotor responsiveness to

cocaine (67, 68). Individually, the Chr17 QTL overlaps with a QTL identified for locomotor responsiveness to nicotine, and the Chr16 QTL overlaps with an interval identified for ethanol preference (206, 207). All of the loci identified in previous research were 50Mbp in size, whereas our support intervals only spanned 2.3Mbp for the Chr17 locus, and 6.3Mbp for the Chr16 locus. Not only does *Lsamp* contain four of the five top candidate SNPs, but it also shows compelling evidence for association with alcohol-related phenotypes in the literature.

Of the three top candidate genes on Chr16, *Lsamp* and *Zbtb20* were especially intriguing. *Zbtb20* encodes a transcription factor that regulates neurogenesis and astrocytogenesis, playing a role in cortical and pituitary development (208-211). It has a particularly well-documented role in the development of the hippocampus (212, 213). These findings are especially relevant, due to the involvement of the hippocampus in motivation, and the well-documented reversible (by alcohol abstinence) reduction in hippocampal volume in long-term alcoholics (214). Additionally, human GWAS, expression, and methylation studies have provided evidence for its association with: risk-taking behavior; seasonality and depression; depression; schizophrenia; seasonal affective disorder; and smoking behavior (170, 215-218).

Even more compellingly, human GWASs have uncovered associations between *Lsamp* and age of onset of AD, as well as several other neurological- and psychiatric-related traits, including: comorbid Major Depressive Disorder and AD, unipolar depression and depressive symptoms, smoking status, mood stability, neuroticism, well-being, feelings of misery, educational attainment, cognition, and comorbid Tourette's Syndrome and Obsessive Compulsive Disorder (170). The protein encoded by this gene (limbic-system-associated

membrane protein) is a neuron surface glycoprotein that contributes to the cell's growth, and to axon guidance and synaptic plasticity in fetal brain and adult limbic system (which is involved in emotional processing and motivation); and is very highly conserved between rodents and humans (214, 219-223). This protein is a member of the immunoglobulin superfamily (IgSF), with high sequence similarity to fellow IgSF family member opioid-binding cell adhesion molecule (221), which is striking because ethanol binds to many opioid receptors. Furthermore, mice deficient in this protein have shown several drug-related, behavioral, and neurological abnormalities, including: decreased sensitivity to amphetamines resulting in increased serotonin release; decreased aggressive behavior; decreased anxiety-like behavior; and decrease in long-term potentiation in the hippocampus (224-226). Given its association with preference for alcohol over water, specifically after long-term alcohol exposure, these results suggest that *Lsamp* may mediate limbic-system neuron synaptic plasticity in response to long-term alcohol exposure, contributing to the development and maintenance of dependence.

Although neither of the top-SNP-containing genes (*Lclat1* and *Gm25406*) shows association with psychiatric and behavioral disorders in GWAS, one of the top candidate genes, *Spast*, has a well-documented relationship with neurological function and disease. Spastin, its encoded protein, targets and cleaves polyglutamated microtubules, and is thereby involved in neurite outgrowth and stability (227-229). Mutations and changes of activity in spastin are associated with paraplegia and Alzheimer's disease, via neurite degenerative processes (230).

Initial Drinking Loci

The support interval of the Chr12 locus for *First Week Mean Total Ethanol Preference* overlaps with loci previously found to be associated with ethanol consumption in female mice and cocaine-induced increase in locomotor activity, which both were 50Mbp wide, compared to our support interval of 1.9Mbp in length (231, 232). *Arg2*, *Plekhh1*, and *Zfyve26* are top candidate genes for this region, and each contains one of the 5 top candidate SNPs. However, the most functionally intriguing of the top candidate genes encodes another microtubule-associated neuronal protein, *Gphn*. Gephyrin is a highly conserved protein responsible for agglomerating glycine and GABA at post-inhibitory-synaptic membranes, thereby increasing signal amplitude and frequency (233) (233). Because of its important neurological regulatory function, mutations in this gene have been found to be associated with autism, schizophrenia, and seizure disorders; and it exhibited associated with unipolar depression in human GWAS (170, 234-236). Interestingly, this protein's function is inhibited by phosphorylation by *GSK3β* (glycogen synthase kinase 3 beta). Expression levels of this gene have been shown to be responsive to acute doses of ethanol in the PFC, and manipulations of its expression levels are associated with changes in voluntary ethanol consumption, such that increases in *GSK3β* expression in the PFC lead to higher consumption (27, 97). Gephyrin is also stabilized by cleavage by calpain-1 (whose gene family member, calpain-13, is a top candidate gene for *Last Week Mean 30% Ethanol Choice*, within our Chr3 locus). Finally, expression levels of a gene encoding GABRG2 (which participates in the gephyrin-mediated post-synaptic clustering of GABA at $GABA_A$ receptors) showed significant expression level differences between postmortem hippocampi of alcohol- and cocaine-dependent individuals and non-addicted controls, and between

alcohol-preferring rats and alcohol-non-preferring rats (237). In light of this compelling evidence for gephyrin's association with alcohol-related traits, we believe that the top-scoring SNPs likely participate in regulating the expression of this *Gphn*, thereby affecting ethanol preference levels.

Another locus was identified for *First Week Mean Total Ethanol Consumption* on Chr6, which overlaps with QTL identified in previous studies for several alcohol- and drug-related phenotypes, including: locomotor responsiveness to cocaine, sensitivity to methamphetamine-induced change in body temperature, ethanol-induced hypothermia, and ethanol-conditioned taste aversion (67, 68, 71, 86). These intervals were all 50Mbp wide, whereas our support interval was 2.57Mbp wide. All of the top 5 SNPs for this region are located in gene *Atp2b2,* one of the top candidate genes for this region, and two (rs30214990 and rs48409474) are classified as noncoding transcript variants. All three of the top candidate genes (*Atp2b2*, *Slc6a1*, and *Slc6a11*) in this region displayed involvement in relevant psychiatric and neurological disorders in the literature. *Atp2b2* encodes a plasma membrane Ca^{2+} pump that appears to play a major role in the neurosignaling involved in hearing and cognition, due to the association of mutations with deafness and autism (238-241). Human GWASs have identified additional associations between cognitive decline, schizophrenia, and opioid dependence (117, 170), with mechanistic and pharmacogenetic studies providing further evidence for its association with schizophrenia (217).

Even more convincingly, human GWASs have indicated that *Slc6a11* is associated with initial alcohol sensitivity, and that the closely related gene *Slc6a1* (also a top candidate gene for this QTL) is associated with longitudinal alcohol consumption, as well as conduct disorder and feelings of loneliness (170). Fittingly, these two genes both encode GABA

transporters, as indicated by their proteins' alternate names: neurotransmitter transporter, GABA, members 1 and 11 (242). Not surprisingly, given their function, other human studies have shown evidence of association of *Slc6a11* with autism, and of *Slc6a1* with schizophrenia, epilepsy, and alcoholism (243-247). Furthermore, a mouse study has shown that GABA transporter 1 antagonists increase ethanol sensitivity (248). Given this evidence, and the known direct effects of alcohol on GABA signaling and this interaction's contribution to dependence (249), we believe it is likely that SNPs in one or both of these genes affected ethanol preference in our study.

Adjacent Chromosome 11 Loci

Both the *Whole Study* and *Last Week Mean Total Ethanol Consumption* QTLs identified on Chr11 overlaps with previously identified QTL for ethanol preference and withdrawal, as well as nicotine-induced changes in locomotor activity, in other mouse strains (114, 206, 250, 251). These intervals were all 50Mbp, whereas we identified support intervals of only 1.56Mbp. Of the top candidate genes for this locus, *Arrb2* possessed the most convincing evidence for alcohol association in the literature. Its encoded protein, β-Arrestin-2, participates in several G-protein-coupled-receptor-related (and some G-protein unrelated) pathways, moderating the responsiveness of these signaling cascades (252, 253). Although it has not shown association with any neurological of psychiatric traits in human GWAS studies, results from several mechanistic studies indicate that it is important to alcohol-targeted signaling pathways. Specifically, lower β-Arrestin-2 levels or receptor responsiveness result in a sensitized reward response to alcohol, lower consumption of and preference for alcohol, and lower Dorsal Root Ganglia opioid tolerance levels in mice (141,

254, 255). Furthermore, like gephyrin, the literature suggests that β-Arrestin-2 interacts with the alcohol-consumption associated gene *GSK3β*, that both of the proteins' expression levels increase in response to long-term alcohol exposure (likely leading to alcohol-induced kidney disease), in an interactive fashion (97, 127, 256, 257).

Lastly, to our knowledge, the only top candidate gene for the Chr11 QTL for *Last Week Mean Total Ethanol Consumption* that has shown association with a neurological, behavioral, or psychiatric phenotype is *Canx*. This gene encodes the protein calnexin, which is a calcium-binding molecular chaperone that interacts with glycoproteins in the endoplasmic reticulum. Although it has not shown association with any neurological or psychiatric traits in GWAS, one study has suggested that mutations in this gene may be associated with familial autism (258). Three of the top five SNPs are classified as intron or upstream variants for *Adamts2*, which encodes a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 2. This protein acts as a protease in the extracellular matrix (259). Although this gene is not in our top candidate gene list, with respect to empirical support from our selected external datasets, other literature provides compelling evidence for the association of this gene with psychiatric and neurological phenotypes. GWASs have found significant associations between this gene and Attention Deficit Disorder, interferon-induced depression, and reaction time (170). Furthermore, its expression levels have been shown to significantly increase in response to alcohol in muscle tissue, and decrease in response to atypical antipsychotics in blood (260, 261). Given its involvement in extracellular matrix structuring, this gene could potentially mediate synaptic remodeling in response to ethanol, by reacting with (and thereby degrading) proteins in the extracellular matrix to change its structure and reroute neuronal processes. In fact, gene

family members *ADAMTS4*, *5*, and *15* have all been shown play a significant role in neuroplastic changes in this way (262).

Summary, Limitations, and Future Directions

In order to obtain substantial power this study used an all-male sample, so that qualitative genetic sex differences would not mask any effects. Therefore, some of the identified loci may not be generalizable to females. However, many of our findings were supported by human GWASs that included both sexes. It is also possible that there are female-specific QTL that could not have been detected in an all-male or mixed sex sample. It is therefore important for future studies to examine these phenotypes in female mice, in order to determine the generalizability of genetic results across sexes, specifically for DO mice.

We anticipated some loci to be identified for the *Whole Study* mean phenotypes that also exhibited lesser (possibly sub-threshold) signals for the *First* and *Last Week* mean phenotypes. These QTL would reflect effects on overall drinking levels, regardless of exposure stage, that would be better captured by phenotypes averaged across all drinking days, reducing noise that could potentially be created by a single aberrant drinking day. We identified one locus for *Whole Week Mean Total Ethanol Preference* (on Chr1) that had a higher LOD score than an overlapping QTL identified for *Last Week Mean Total Ethanol Preference*, and had an overlapping sub-threshold peak for *First Week Total Ethanol Consumption*. This region contained five genes that showed significant association with "drinks per week" in the GSCAN GWAS sample (61). Upon further investigation, we found that these genes are also adjacent to one another on Chr1 in humans, indicating that this genomic region is conserved between mice and humans (263). A candidate gene for on other

QTL, the Chr6 locus for *First Week Mean Total Ethanol Consumption*, *Slc6a1*, has been found to be significantly associated with ethanol consumption in a human GWAS sample (170). Another gene in that interval, *Slc6a11* (in the same gene family as *Slc6a1*) was identified with initial alcohol sensitivity in human GWAS, along with *Tm4sf5*, a candidate gene on the Chr11 locus that was uniquely identified for *Whole Study Mean Total Ethanol Consumption* (170). No candidate genes from QTL that were unique to Last Week drinking behaviors have been identified for either of these phenotypes in human GWAS. However, all three of the candidate genes from this study that have been identified in AD GWASs (*Impad1*, *Lsamp*, and *Ubl4b*) were associated with QTL for *Last Week* phenotypes.

Finally, only one QTL was identified for 30% ethanol choice, and it was only identified for the last week time interval. Because no taste additives were present in the alcohol solutions, higher levels of preference for 30% ethanol indicate that, despite the more noxious taste, mice prefer the heightened level of psychotropic reward provided by 30% ethanol over 15% ethanol. These findings suggest that our hypothesis was correct, not only in that we would identify some common and some unique QTL between time intervals, but also in that the first week of drinking appears to model initial sensitivity and non-pathological drinking in humans, whereas the last week is modeling a phenotype that better models dependence. The two QTL that appeared to represent genetic effects that were relatively consistent across time appeared to be modeling genetically driven alcohol consumption and sensitivity in humans, like those identified for initial drinking behaviors.

In sum, this study identified one novel locus for ethanol consumption and preference with greater precision, and several previously identified drug- and ethanol-associated loci, with far greater precision than prior studies. We were also able to gain insight into the

differing biological mechanisms that underlie initial drinking behaviors and those that occur after long-term ethanol exposure. Our results indicated that genetic pathways associated with voluntary ethanol consumption behaviors under an intermittent ethanol access paradigm qualitatively differ between initial exposure and post-long-term exposure time points, as was expected for many of the identified variants. They also suggest that initial voluntary ethanol consumption models ethanol sensitivity and consumption in humans, whereas long-term consumption under IEA models mechanisms relevant to dependence. Specifically, initial drinking appears to be associated with pathways associated with GABA-related signaling, based on the function of top candidate genes *Slc6a1*, *Slc6a11*, and *Gphn*; whereas long-term drinking genetic effects appeared to be mediated by pathways related to neurogenesis and neurite outgrowth, immune and neuroinflammatory responses, and presynaptic vesicle anchoring. The relevance of these pathways to alcoholism is well supported by studies displaying the neuroinflammation seen in the brains of chronic alcoholics, and neuroplastic changes that are involved in the development of dependence (9, 15, 264, 265). Future studies should take this into consideration when studying voluntary ethanol consumption in mouse models, as it appears that long-term consumption may be a better biological model dependence-like drinking and initial consumption may be a better model for reward-associated learning and sensitivity to alcohol with respect to physical symptoms of intoxication. Such phenotypes should be examined in mice in which one of a candidate gene identified in this study is over- or under-expressed, to determine direct effects of each gene on drinking behaviors. One potential way to narrow down the candidate gene list further, based on our candidate SNP findings, would be to use Capture-C to determine which regions of the chromosome are physically interacting with the SNP-containing region (266). Another

option would be to mutate each of the candidate SNPs in an inbred mouse strain using CRISPR-Cas9 technology, and examine its effect on both behavior and on gene expression levels, in order to determine which genes each SNP regulates and which SNPs are the true causative SNPs. Overall, our study identified several candidate SNPs and genes that effect distinct periods of voluntary ethanol consumption in the Diversity Outbred mice. These genes and variants can be targeted in future functional studies in order to determine the exact biological mechanisms involved in these behaviors that can be targeted by pharmacotherapies for alcohol dependence.

Chapter 5

Characterization of Genome-Wide Gene Expression in the Prefrontal Cortex

Introduction

As described previously, human genetic association studies for alcohol use disorders have had relatively limited success in accounting for the total heritability of these traits estimated by twin studies (~50%) (21-23). Recent studies with hundreds of thousands participants have come closer to this goal, identifying up to 90 variants associated for ethanol consumption, due to increased power provided by the large sample sizes (25, 26, 61). However, even after significant genetic associations have been identified, many of these results do not replicate across samples, there is no mechanistic explanation for the association beyond that which can be gathered from unrelated, previously performed, functional studies. Gene expression networks are capable of providing information regarding groups of co-functioning or co-regulated genes whose overall biological function and mechanisms of regulation are related to the trait of interest. Our findings in *Chapter 1* suggest that different genetic associations found in separate populations may be identifying similar mechanistic pathways. This suggests that gene networks not only provide information regarding functional networks through which these genes may affect the examined trait, but also provide testable groups of genes that may be more powerful for association analysis than any single constituent gene, due to combined effects. Furthermore, although some studies have identified differentially expressed genes in brain tissue between alcoholics and that from non-alcoholics, gene expression data collected from post-mortem brain tissue samples can be unreliable and difficult to interpret. This is due to lack of experimental control and limited sample sizes. For these reasons, many researchers have turned to mouse models to examine gene expression networks in the brain relevant to physiological response to alcohol and alcohol-related behaviors (28 2012, 96).

Several gene expression studies have successfully identified networks of co-expressed genes that are differentially expressed between alcohol-exposed mice and controls in brain regions related to the dopaminergic reward pathway (specifically, the Ventral Tegmental Area, Nucleus Accumbens, and Prefrontal Cortex (PFC)) (27-30). Some of these studies identified a network that was overrepresented with human GWAS signals for alcohol dependence, one of which possessed a hub gene that has since been shown to be a promising target for pharmacotherapy to treat alcoholism (27, 29, 97). The study presented in Chapter 1 of this thesis examined gene networks whose structures were perturbed by alcohol treatment in these three brain regions, in combination with human alcohol dependence GWAS summary statistics and protein-protein interaction data. To identify networks associated with alcohol consumption behaviors, this chapter examines gene expression networks in the prefrontal cortex of the voluntarily drinking Diversity Outbred mouse sample described in the previous chapter. The prefrontal cortex was chosen due to its involvement in the dopaminergic reward pathway, but also because of its role in behavioral inhibition and decision-making. The Intermittent Ethanol Access (IEA) paradigm to which these mice were exposed models the progressive increase in alcohol consumption seen in the early stages of dependence development, despite the noxious taste of ethanol and the unpleasantness of intoxication (36, 37, 267).

Specifically, we map networks of co-expressed genes in the PFC of 100 of the highest drinking and 100 of the lowest drinking IEA-exposed DO mice. We then test for overlap between bQTL identified for drinking behaviors in the previous chapter and expression QTL for each network's first principal component (representing the overall variance in expression throughout the network). We also test for overrepresentation of

networks with functional groups and with genes significantly associated with ethanol consumption in the Genome-Wide Association Study (GWAS) Sequencing Consortium for Alcohol of Nicotine use sample (GSCAN). Finally, we examine individual genes whose expression levels significantly predict ethanol drinking behaviors, and genes whose expression levels differ between high drinking DO mice and controls (exposed to water only). Genes exhibiting significance in both analyses are considered to be ethanol-regulated. However, as described in Bogenpohl et al. (in progress), a study that analyzed gene expression in tissue from chronically voluntarily drinking Rhesus Macaques, we anticipate there to be genes that will not differ between drinkers and controls in DEseq, but will have a continuous relationship with drinking values. Although causation cannot be confirmed in this study, such genes are thought to be ethanol behavior-regulating, as opposed to alcohol-regulated. This means their expression levels do not change in response to ethanol exposure, but expression differences between mice drives variation in alcohol consumption and preference levels (Bogenpohl et al., in progress). We may also see genes that have a significant continuous relationship with ethanol drinking behaviors but do not show significant differential expression between IEA-treated and control mice, which was also reported by Bogenpohl et al. (in progress). Such results will be interpreted as follows: there is likely a ceiling effect of alcohol on the gene's expression levels that is reached at relatively low doses of ethanol, resulting in a true relationship that appears to be dose-independent. Whole networks were also tested for expression levels that correlated with ethanol consumption and for overrepresentation with genes that are differentially expressed between high drinkers and controls, in order to determine which networks are likely ethanol-regulating vs. ethanol-regulated. Just as for the single gene analyses, we anticipated the

identification of several networks that were ethanol-regulated and some that were ethanol-regulating (i.e. that moderated ethanol consumption behavior). Overall, this study was the first to successfully identified such modules in a Diversity Outbred mouse sample. These networks were functionally overrepresented for pathways largely involved in neurocognitive development, synaptic signaling, protein modulation and regulation, and the Notch signaling pathway.

Methods

Experimental Design

This experiment utilized the same sample of 636 adult male Diversity Outbred mice reported in *Chapter 4*. As previously described, these mice were exposed to an intermittent ethanol access (IEA) paradigm, in which mice were subject to alternate periods of access to ethanol and water, and periods of access to water only (forced ethanol abstinence). The experiment followed a three-bottle choice paradigm, in which mice were given access to one bottle of 30% EtOH, one of 15% EtOH, and one of water on ethanol access days, and to three bottles of water on days of forced abstinence. Ethanol access was granted for 24hr periods, 12hrs in the light and 12hrs in the dark, 3 days per week for 4 weeks. Controls (N=49) had access to water only, throughout the course of the experiment. Mice then underwent post-anxiety-like behavioral testing and continued to drink for one more week, in order to re-establish ethanol-driven expression patterns in the absence of additional testing. Humane sacrifice, and immediately subsequent tissue and trunk blood collection, occurred 24hrs

following their last ethanol access period. Brain tissue was microdissected into 8 regions, one of them being the Prefrontal Cortex for gene expression analysis.

Gene Expression Data Collection and Assay

Mice were chosen for RNAseq by their average total ethanol consumption values during the last week of drinking, before anxiety-like testing began (IEA week 4), with 100 mice being chosen from each extreme of the distribution. In addition to these 100 high drinkers and 100 low drinkers, tissue from 20 controls (exposed to water only) was also sequenced. RNA was extracted from Prefrontal Cortex tissue using Qiagen RNAeasy mRNA extraction kit. Samples were randomized, prior to extraction, to prevent correspondence between extraction variability and mouse cohort. RNA purity and concentration were examined using a Thermo Scientific NanoDrop. All samples had 206/280 values ≥ 1.9, indicating high RNA purity. RNA samples were shipped on dry ice to Novogene Bioinformatics Institute, Sacramento, California where poly-A selected library construction and sequencing were performed. Before sequencing, the quality of each sample was checked via the 2100 Agilent Bioanalyzer. All samples had RNA Integrity Numbers (RINs) ≥ 7, with 93% of the samples having RINs ≥ 9. Only 2 samples yielded < 1µg of RNA, the lowest yield being 0.46µg. However, these samples had good RIN numbers (9.8 and 9.3). Therefore, these two samples were moved forward to sequencing. The Illumina HiSeqX was used to pair-end sequence 150bp fragments, with 20 million reads per sample. All samples had > 90% alignment, with an average of 94.46% (Table 1). Overall alignment percentages are expected to be slightly lower than what would typically be seen (in the high 90s), due to the use of a single inbred mouse genome as the reference to align samples that possess up to 45mill variants that differ from this reference (171). However, algorithms and statistical

packages for alignment to personalized genomes are still being perfected with respect to alignment percentages, and are only more accurate than standard alignment when examining novel variants or computing allele specific expression (171, 268). A minimum of 95.14% and 88.73% of reads in each sample had Phred Quality scores of Q20 and Q30 or higher, respectively, indicating high read quality. The minimum number of flow cells (241) was used, and were run on the same machine, to reduce technical variability. A minimum of 95.14% and 88.73% of reads in each sample had Phred Quality scores of Q20 and Q30 or higher, respectively, indicating high read quality. Samples that did not reach a minimum of 18.5 million raw reads (N=9) were re-sequenced up to 20 million reads.

Alignment and Read Count Processing

Post-sequencing quality control was performed using FastQC 0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality was assessed based on several metrics: base sequence quality, tile sequence quality, base sequence content, GC content, sequence length (should equal 150bp), sequence duplication levels, and adapter content. As typical of RNAseq analysis, base quality values at the far ends of each read were lower than the rest of the fragment (Figure 1). Therefore, Trimmomatic 0.38 (269) was used the trim the first 5bp from the 3' end of each read, and trim the 5' end based on quality (sliding window method). FastQC was performed again after trimming, and all samples passed QC. Although there appeared to be slight differences in per-sample average sequence quality between flow cell lanes, all lanes yielded high-quality reads (Figure 1).

The RNAseq processing pipeline is laid out in . Samples were aligned to a C57 reference genome (GenCode, Assembly GRCm38.p6) with STAR 2.6.1a, using default

settings for advanced alignment options (270). All samples had > 90% alignment. The alignment rates were expected to be lower than those that would be typical for inbred strains, given the large number of SNPs in the DO genomes compared to the reference genome. Gene counts were determined by HTSeq 0.8.0, using the "union" method to handle multi-mapping reads, and using an alignment quality threshold of 10 (i.e. removed reads if alignment quality < 10) (271). After quantification, all ribosomal and mitochondrial genes were removed. These values were converted to Transcripts per Million (TPM) before genes with extremely low expression values were removed, due to the likelihood that these signals were attributable to noise. Genes were removed if > 66% of the TPM values were ≤ 0.5 (https://github.com/mdozmorov), leaving 17,709 genes for analysis. Failing genes were also removed from raw gene count data, before performing DEseq2. To satisfy the normality assumptions of correlative and regressive methods, we used the $\log_2(1+TPM)$ as expression values for all analyses, except DEseq2 (which requires raw, untransformed count data) and WGCNA (which requires batch corrections, described below).

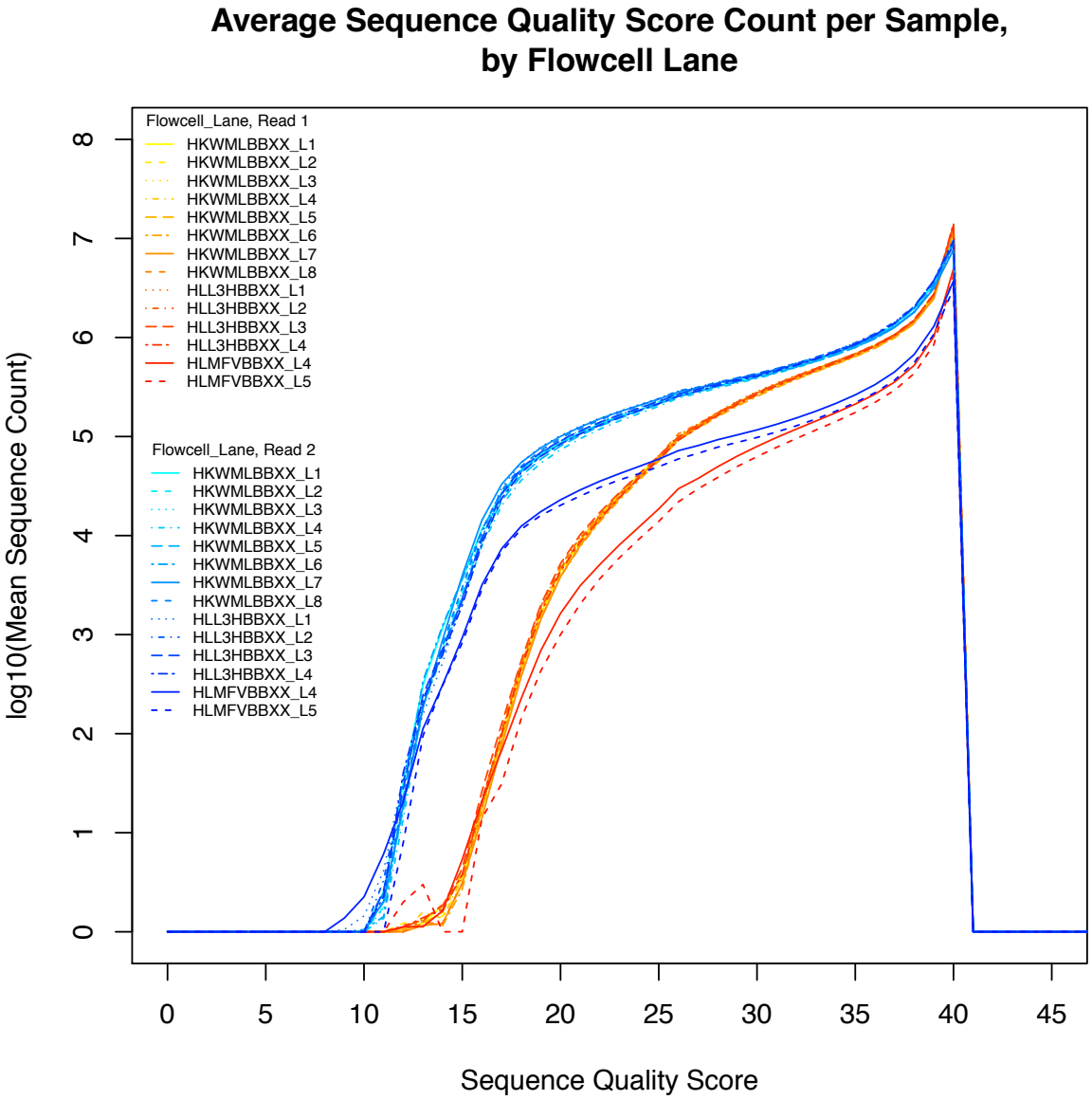
Table 1: Alignment Percentages

Sample	Total.Counts	Unmapped	% Aligned
DO100	60061542	2805612	95.32
DO107	59984080	3296966	93.91
DO108	54127782	2591640	95.21
DO109	54131948	7373076	90.90
DO11	80991436	3912004	94.51
DO110	71306000	3299538	93.99
DO114	54924774	3095858	94.77
DO116	59213674	2541140	96.34
DO118	69337118	5631120	90.15
DO120	57172250	3005066	94.39
DO122	53536656	2967544	94.88
DO123	57961234	2786326	95.49
DO125	61759062	3171606	94.82
DO126	61179730	2843724	94.79
DO128	54574032	3366570	93.69
DO130	53339298	3029026	95.27
DO132	64050344	3275500	94.28
DO133	57221488	3151306	94.66
DO134	59060932	3146202	94.85
DO135	61124944	3776386	94.34
DO141	66750010	3533978	94.62
DO143	65722422	5146336	91.03
DO144	57393674	3610910	93.67
DO146	57023630	3375814	94.43
DO149	60569412	3305740	95.58
DO154	74854534	4157218	92.51
DO156	55535408	2939792	95.16
DO159	60781302	2794082	95.40
DO165	65549792	3731320	94.40
DO171	66617802	4038710	93.23
DO174	59672922	2676650	95.64
DO175	61400400	3324374	95.40
DO177	72300172	3916886	93.38
DO179	59196528	2976534	95.39
DO187	64588844	3424166	94.48
DO189	62066372	3488748	93.98
DO190	57977934	2748566	95.52
DO194	61361530	3360528	94.17
DO196	57635288	3089116	95.19
DO197	64187860	3218052	94.91
DO200	63231204	4935442	91.75
DO204	59815016	2778566	96.49
DO205	79271054	4627594	92.71
DO207	63491078	3493982	94.07
DO208	58877904	5029220	93.20
DO209	74010456	4154802	93.27
DO210	61728946	3118768	95.39
DO211	67652216	3672048	94.33
DO212	64812564	3171132	95.33
DO215	67966682	3763628	94.62
DO221	69916578	4001994	93.38
DO222	60428146	3274972	95.31
DO232	69775884	5055824	90.99
DO233	56112512	3563204	94.05
DO238	59858906	3246276	93.86
DO239	52886702	2773932	95.17
DO24	57375224	2478728	95.17
DO241	51279044	2731298	95.83
DO244	65486004	3734788	93.38
DO246	56405998	3823172	93.90
DO247	62687214	3423190	94.41
DO249	61267248	3132916	93.76
DO255	50218496	2591260	95.66
DO256	59679234	3279772	93.77
DO262	52611434	4578528	92.51
DO264	61137734	3416214	94.49
DO270	62013018	2867062	95.18
DO278	59486302	2617486	95.53
DO279	58549750	3332402	93.77
DO280	53506708	2739266	95.12
DO285	56098488	3095140	95.01
DO289	61981960	3222024	94.14
DO29	54995770	2698648	96.08

Sample	Total.Counts	Unmapped	% Aligned
DO290	68854424	3925048	94.51
DO291	71472948	3991576	93.37
DO294	60185186	3015752	96.09
DO295	77037432	3795106	94.59
DO297	70170606	3715916	94.20
DO298	64097424	3455950	94.69
DO30	65092922	3026106	94.59
DO300	55920042	3130716	94.23
DO306	54302532	3001040	94.65
DO307	56064850	2578042	95.20
DO309	53734954	2963886	94.55
DO311	54412274	2832794	96.14
DO312	73358706	3957524	92.94
DO320	56033856	2508618	95.09
DO321	51101002	2502874	95.78
DO328	59286006	3134116	95.21
DO329	65444390	3355354	94.32
DO333	59037682	2546494	96.46
DO334	71975510	3838442	93.45
DO337	58636806	4354444	92.45
DO339	57685196	3310020	95.42
DO34	72328256	3678740	93.76
DO349	58953030	5052100	91.56
DO350	59863926	3232960	94.78
DO352	61935954	3398832	94.52
DO354	62027456	3155858	94.27
DO357	55028106	2919582	94.55
DO363	53521594	3287566	94.47
DO370	59427882	2717600	95.15
DO374	56003660	3065636	94.59
DO38	56665730	2980490	94.59
DO382	55095516	2771136	95.25
DO383	58329854	3462276	93.38
DO384	52334428	2835584	94.50
DO390	51531008	2854266	94.96
DO395	56621774	2979194	95.02
DO401	59767234	3275800	94.34
DO404	57901000	4650882	92.18
DO405	59471608	3400754	94.33
DO406	60011210	3526338	94.92
DO410	69471166	3479948	94.44
DO412	62582952	3415702	94.49
DO415	61963898	3205404	94.33
DO416	56527980	3059160	95.36
DO42	65918402	4131914	93.38
DO420	62371430	3485238	94.41
DO435	67764034	3328222	94.65
DO439	62213940	3284706	94.72
DO440	54189998	3307210	94.39
DO441	58956512	4748108	91.94
DO443	58938274	3582732	93.45
DO448	54686876	2651720	95.63
DO45	60728150	3613058	94.14
DO450	61681994	3376430	94.80
DO451	64958862	2912274	96.23
DO453	77328534	4573738	92.14
DO454	58158262	2813364	95.80
DO457	67033182	3627668	94.59
DO464	65196384	3124062	94.34
DO47	55238862	3283724	94.06
DO480	55637752	2965520	94.57
DO481	54591708	4681558	92.43
DO485	61860702	3555776	94.20
DO488	61286972	3490376	94.41
DO490	62444002	3159186	94.89
DO493	61825072	3517398	94.26
DO494	61229264	2931280	96.19
DO496	76937110	3776022	93.65
DO498	59477566	3258660	94.08
DO509	55027588	2913696	95.19
DO513	60620452	4355208	92.82
DO519	64683066	3038750	95.04
DO52	61312456	2883306	95.69

Sample	Total.Counts	Unmapped	% Aligned
DO520	66949192	7450484	88.06
DO521	62406630	3862764	93.22
DO522	57003326	3363450	94.20
DO523	58001646	3319378	95.56
DO526	74747588	3817808	94.04
DO530	64004398	3011390	95.30
DO534	66147368	3900570	94.06
DO537	65674894	4416276	91.20
DO545	50197412	2825816	95.85
DO550	68030698	3420632	94.30
DO552	60019622	2892246	95.14
DO554	59552996	2928150	95.14
DO558	60282852	3147422	95.12
DO56	64431636	4446434	93.56
DO560	69020852	4154952	93.25
DO563	61581426	3439106	94.10
DO564	58283936	3388518	94.01
DO569	56556464	3283632	93.93
DO57	54103040	3437792	94.20
DO571	59255864	2512492	96.07
DO572	63972442	3307326	94.21
DO575	57116544	2767572	96.19
DO577	72728350	4268242	92.93
DO579	60329660	3011226	94.78
DO580	57631976	3904222	94.31
DO581	68568906	3681168	94.63
DO588	63027572	3990602	92.49
DO589	53113592	2822590	95.11
DO59	57718814	2951894	95.71
DO590	68876472	4276852	93.53
DO592	66089236	3881934	93.40
DO596	58854248	3120448	94.44
DO597	56094710	3155744	94.29
DO599	55292604	3235742	95.10
DO60	66014984	3583952	94.50
DO600	65175416	3622230	94.70
DO601	68400126	3484454	93.90
DO604	57080504	3324392	94.91
DO606	65321346	3931814	94.15
DO609	67200726	3615088	94.58
DO61	66645850	4266034	92.86
DO616	59740004	3384656	93.69
DO619	53633082	2704274	96.49
DO621	77028670	4520022	91.42
DO625	52658372	2407212	95.91
DO629	58897234	3212504	94.90
DO64	62998294	3393738	95.59
DO68	76868902	3800362	93.87
DO7	62041636	4434286	92.03
DO71	55612608	2962358	95.13
DO74	60849464	2550626	95.79
DO75	60654222	3636290	93.72
DO76	57906114	3317224	93.49
DO77	50986462	2570020	96.36
DO78	70562482	3008610	94.81
DO8	57951766	4087380	94.14
DO80	69700998	3604044	94.12
DO81	61329568	3106710	93.97
DO82	51519238	2689808	95.76
DO84	63502192	2712790	94.93
DO86	53519968	2830042	94.86
DO92	55103050	2852168	96.06
DO94	72350666	2757940	95.35
DO96	59312542	3473256	94.23
DO98	60147052	2781666	93.03
DO105	39923306	1212683	96.82
DO164	38133769	1194554	96.63
DO434	35431831	1265280	96.69
DO44	38202609	1061161	97.26
DO458	38700376	1219663	96.64
DO474	36277231	1053105	97.18
DO514	37367937	1175353	97.97
DO532	57798447	2059952	96.44
DO587	61967046	3042650	95.09

Total number of RNAseq read counts, total number of counts aligned, and alignment percentages for each sample.



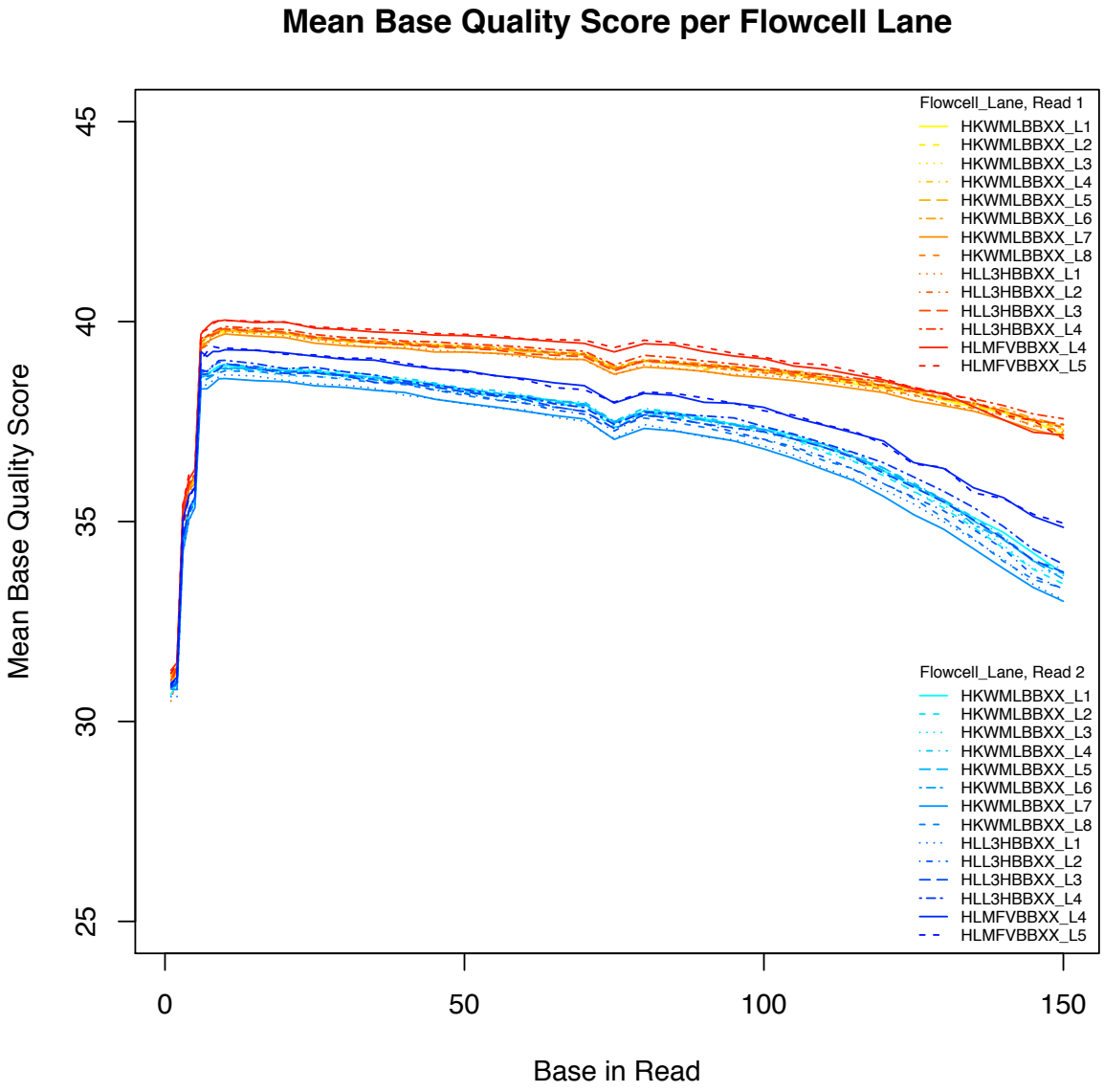


Figure 1: Average sequence quality score count (a) and base quality score count (b) across the length of reads, for each flow cell lane. Colors and line type represent a single lane on a single flow cell, as shown in the legends. Red hued lines are for the first read, of paired end sequencing, and blue hued lines are for the second read. First reads are expected to have higher values (notice log scale in (a)), as depicted here. Overall, flow cell and lane did not have any striking effects on read quality.

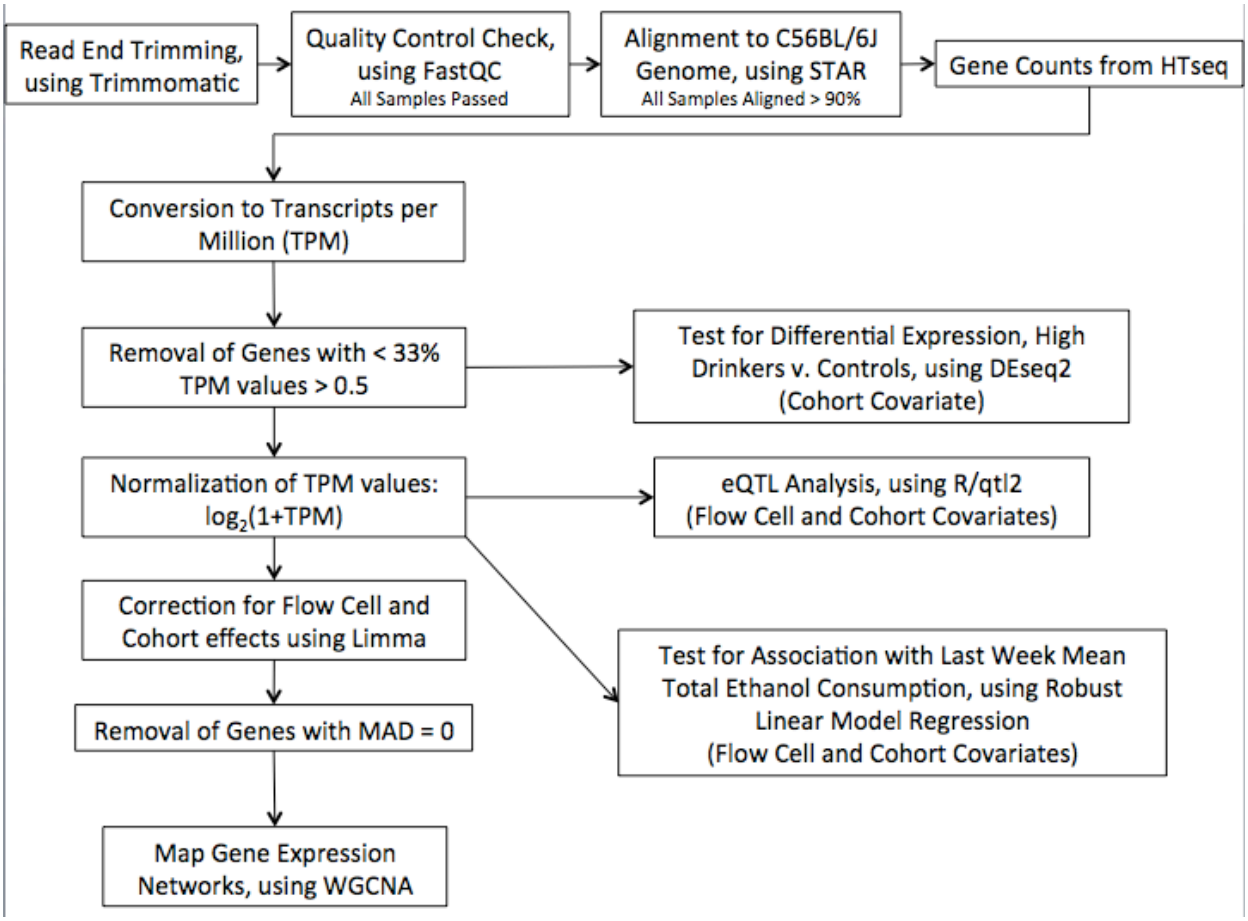


Figure 2: RNAseq data processing pipeline for each analysis described in this chapter.

Statistical Methods

Principal component analysis was performed on log-transformed TPM values, in order to examine effects of *Cohort* (which directly correlates with DO mouse generation) and *Flow Cell* on expression values, because experimental groups and drinking groups were not distributed evenly across cohorts (Figure 3). There was clear clustering by *Flow Cell*, but the effects of *Cohort* on expression were less clear (Figure 4a-b). Because low drinkers and controls on the first flow cell cluster with the high drinkers on the same flow cell, as opposed to clustering with the low drinkers on the second flow cell, it is clear that these effects are largely due to technical differences across flow cells and not experimental differences (drinking amount) between groups (Figure 4c). For clarification, we examined the significance of effects of *Cohort* on the first three principal components (PCs) via logistic regression, while including *Flow cell* as a covariate, to remove technical variation. Cohort did significantly affect the second and third principal components. For our network analysis, which does not allow for the inclusion of covariates, we used the Limma R-package to remove the effects of *Flow Cell* and *Cohort* from log-transformed TPM values (272). For all other analyses, *Flow Cell* and *Cohort* were included as covariates while analyzing log-transformed TPM values. For all analyses, for samples that required data from a second round of sequencing, *Flow Cell* was determined by the original *Flow Cell* on which that sample was run, because these samples clustered well with other samples from their original flow cells (Figure 3), as the majority of reads for these samples came from the original flow cells.

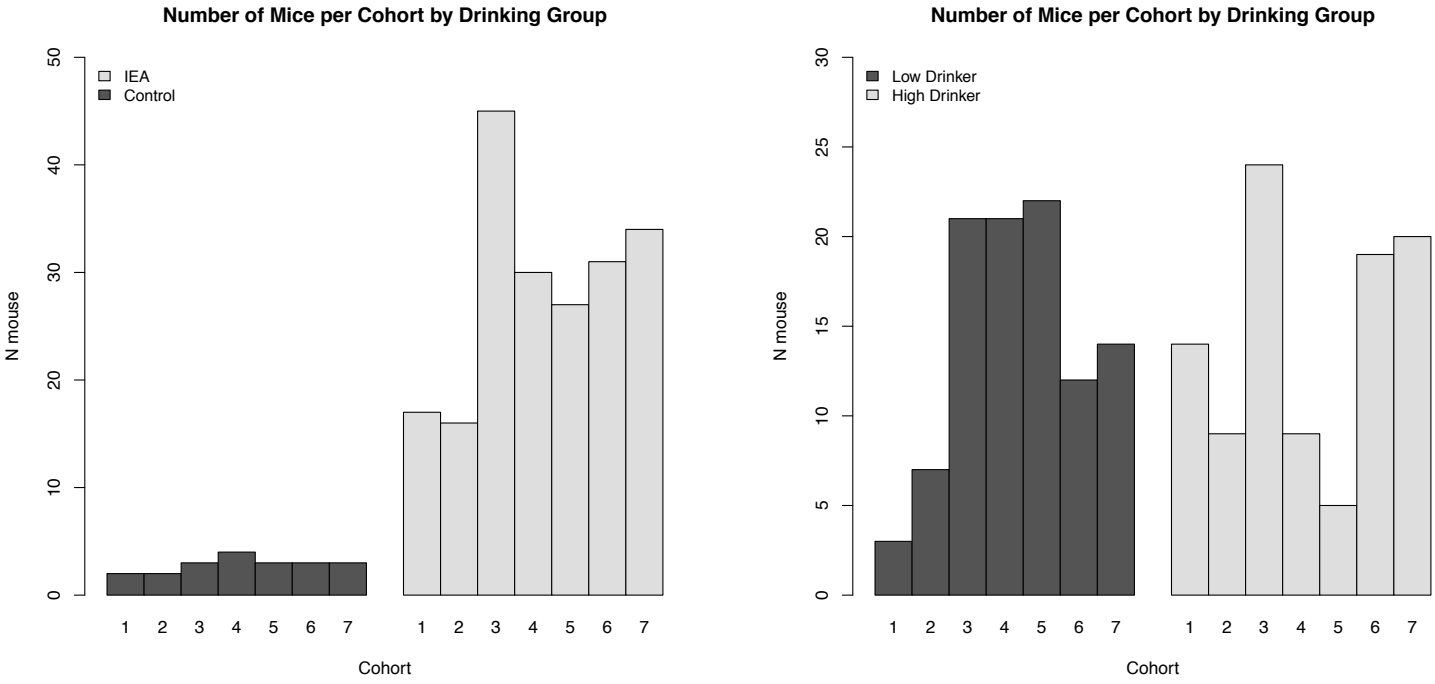
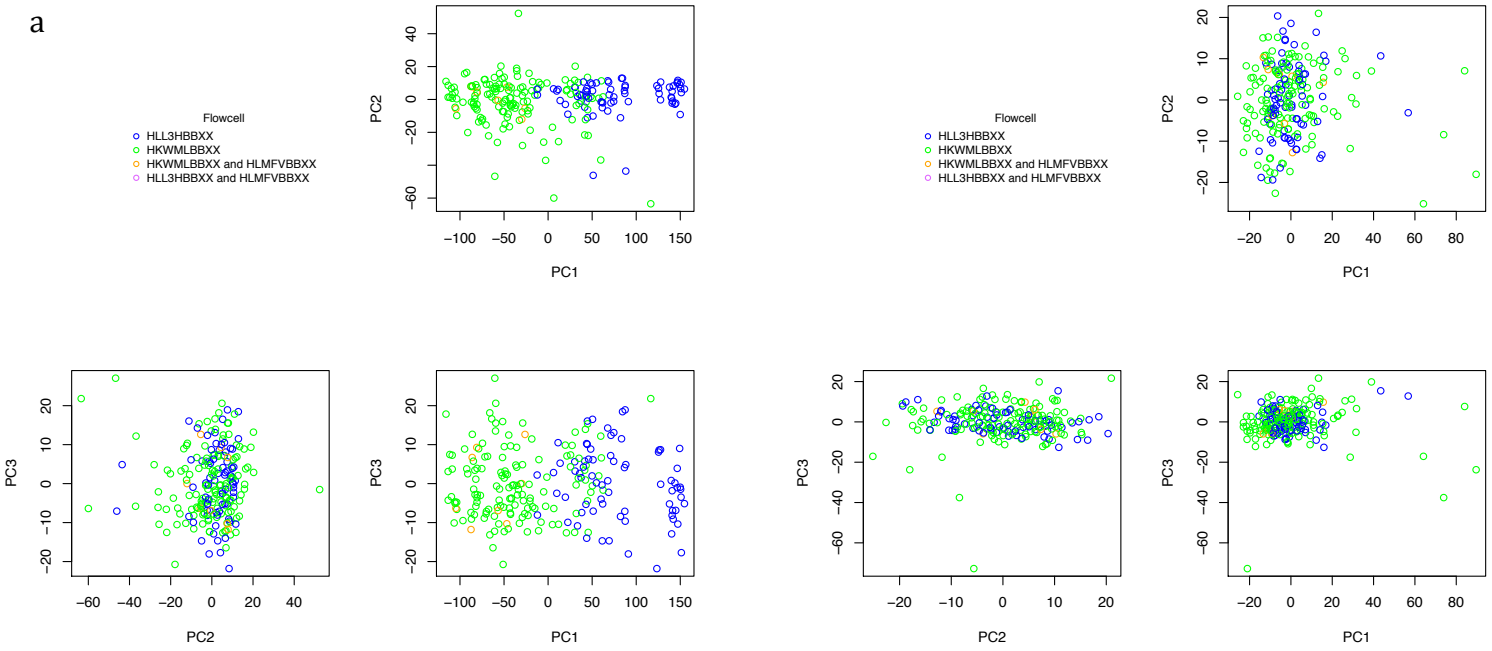
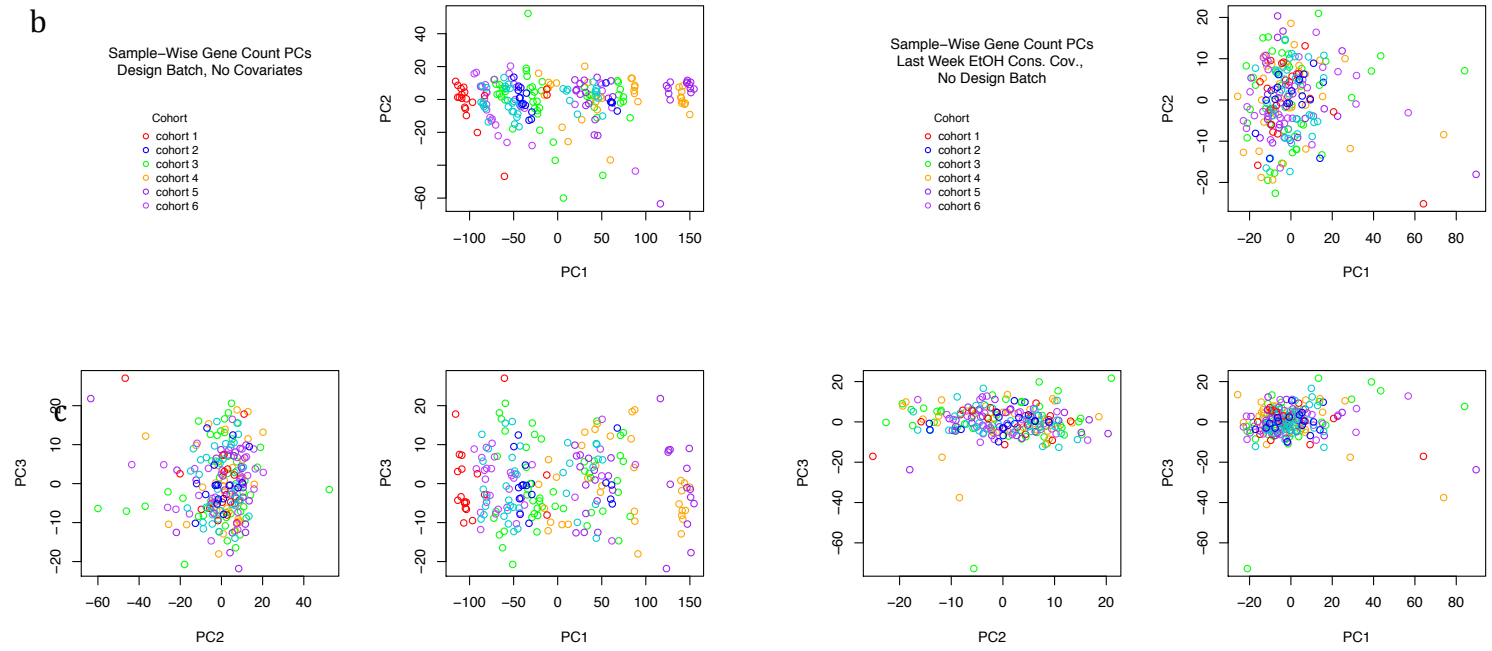


Figure 3: Histograms of mice per cohort, separated by experimental group (left) and drinking group within the ethanol-exposed experimental group (right).

a



b



c

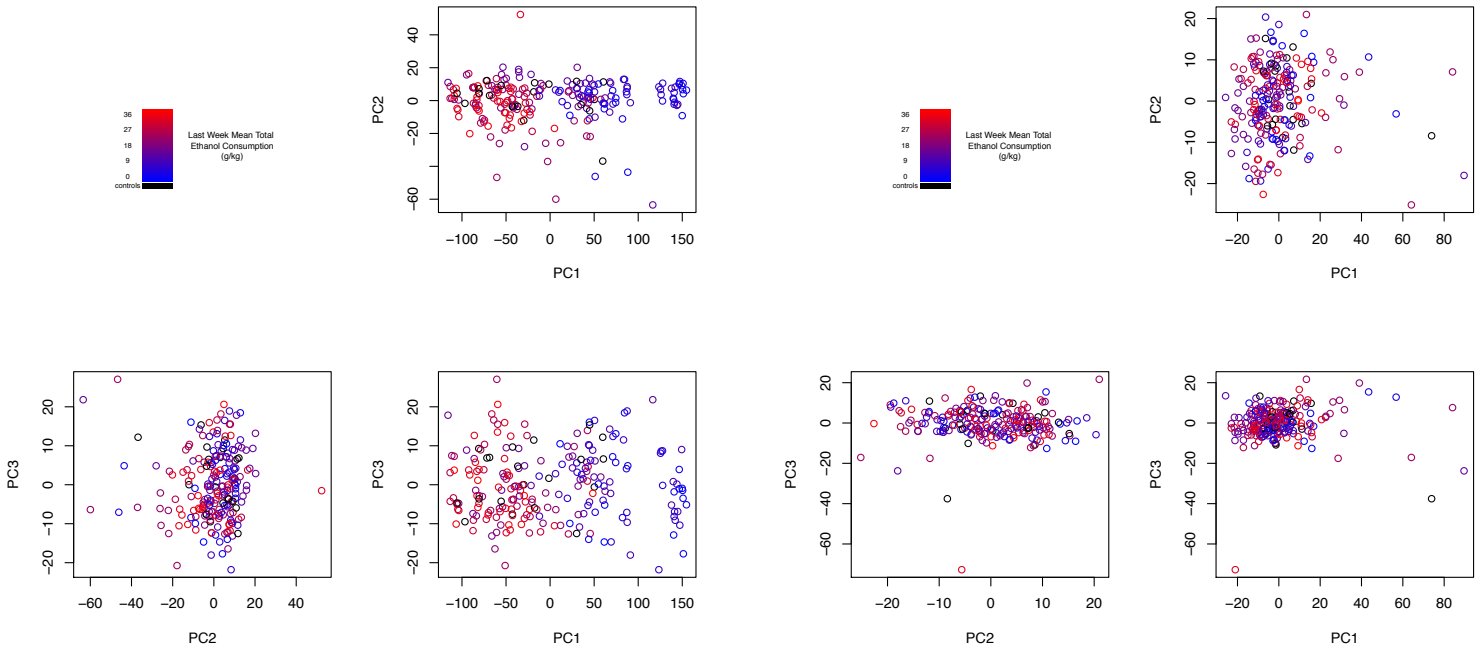


Figure 4: Gene expression principal component plots for samples, colored by RNAseq flow cell (a), Experimental Cohort (b), and Drinking Level (c), before (left) and after (right) Limma correction for Flow Cell and Experimental Cohort. Lack of clustering in the right-hand plots indicates that Limma successfully corrected for covariate effects, and that no other covariates with strikingly visible results exist.

Differential Expression and Expression-Phenotype Associations

Differential expression between the highest drinkers and controls was tested for each gene using DEseq2. The lowest drinkers were not included in this analysis, as many of them consumed negligible amounts of ethanol, which means their ethanol exposure levels were similar to that of controls. Because the goal of this particular analysis was to detect ethanol-regulated genes (i.e. genes whose expression levels change in response to ethanol), as opposed to ethanol-regulating genes (i.e. genes whose expression levels moderate ethanol consumption levels), the inclusion of the lowest drinkers in either group would reduce our power. Furthermore, correction for flow cell differences would decrease power to detect true biological differences between low drinkers and high drinkers, because the majority of the low drinkers were placed on the same flow cell. The comparison of high drinkers v. controls

did not require this power-limiting correction, because they were sequenced on the same flow cell. The program utilizes a Wald test to examine expression differences between two groups. It adjusts p-values using the Benjamini-Hochberg formula, after removing values for genes that did not have sufficient variance for the implemented normalization algorithm (273). As mentioned above, this analysis was performed on raw gene counts, and *Cohort* was included as a covariate. An adjusted p-value < 0.1 was considered to represent significant expression differences between groups. To take advantage of the continuous nature of our alcohol consumption phenotype, we also tested for a relationship between expression values and *Last Week Mean Total Ethanol Consumption*. Unlike DEseq2, this analysis examined the linear relationship between continuous drinking values, as reflection of a behavioral trait as well as biological exposure. Therefore, high and low drinkers were included in this analysis, and flow cell was incorporated into the regression model as a covariate. We used Robust Linear Model regression from the MASS R-package to test for linear relationships between log-transformed TPM values and the following phenotypes, using expression as the predictor and ethanol consumption as the dependent variable: *Last Week*, *First Week*, and *Whole Study Mean Total Ethanol Consumption* and *Preference*. This method was used because it is robust against non-normality of dependent variables, as is seen in our drinking data, due to our sampling of extreme values (100 highest drinkers and 100 lowest drinkers). As for other analyses, Cohort and Flow Cell were included as categorical covariates. Significance was defined as a False Discovery Rate (q) < 0.1, calculated within each drinking phenotype. Because high drinking group in the differential expression analysis was designated based on last week mean total ethanol consumption levels, differential expression results will only be

compared to results from the linear regression testing for expression association with last week mean total ethanol consumption, for consistency.

Gene Expression Network Analysis

We used Weighted Gene Co-expression Network Analysis (38) to map scale-free networks, or "modules", of genes with highly correlated expression values (38). This method works via hierarchical clustering, based on topological overlap, which takes into account adjacency (or transformed expression correlation) and network topology (or the connectedness of two genes, based on their shared connections). Because this method does not use algorithms that allow for the inclusion of covariates, expression data was corrected for effects of RNAseq *Flow Cell* ($N_{SamplesFC1}$=148, $N_{SamplesFC2}$=42) and *Cohort* (groups in which mice underwent experimentation) before network construction, using R-package Limma's batch effects correction function (272).

WGCNA is a weighted network-mapping method meaning there is no hard association threshold for the presence of a connection between two genes to be granted. The resulting networks are scale-free, meaning the probability of being connected to a certain number of nodes follows a power distribution (100). Scale-free networks are characterized by the presence of highly connected "hubs", rendering robustness of the network structure against the removal of single edges (or connections between genes) (100). These types of networks are commonly occurring in nature and are a good representation of biological frameworks, for which there are often molecular "hubs" that affect many other peripheral genes or proteins (100). In order to achieve optimal scale-free nature, while balancing the connectivity (a value that incorporates the number and strength of connections between one gene and all other genes), the adjacency values must raised to a user-specified power.

Typically, scale-free conformance increases and connectivity (for which moderate values provide the most informative results) decreases, both asymptotically, as this power value increases in pre-clustered (by K-means clustering) networks. However, because this relationship between soft threshold value and scale-free conformance was displayed by our data, we set the power to which the adjacency values would be raised to 6, as suggested in Horvath et al., (100) (Figure 5). We used the unsigned bi-weight correlation method (which is robust to outliers) for adjacency calculations, and determined module membership by a dendrogram (or relatedness tree) with a deep split cut parameter of 4, in order to maximize sensitivity, allowing modules to be no smaller than 30 genes. Modules were merged if their branches met at a dendrogram height < 0.05. These networks were analyzed in ethanol-exposed mice only, as ethanol can affect correlations between gene expression values, and therefore network structure (30, 274).

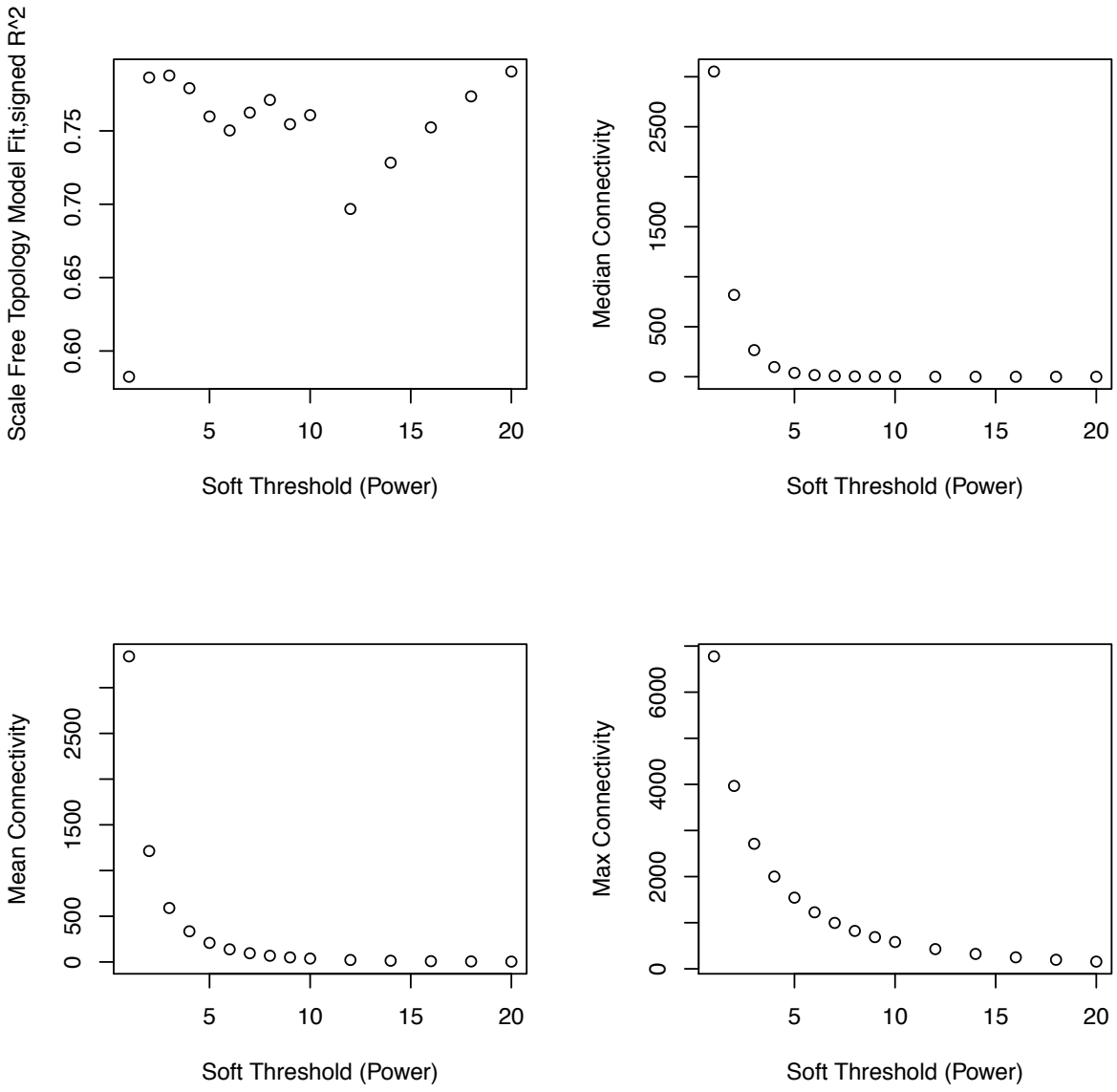


Figure 5: Scale-free topology fit index, and mean, median, and maximum connectivity for rudimentary modules mapped via K-means clustering, for each of several potential adjacency power thresholds. The ideal power value maximizes scale-free fit while keeping connectivity at moderate, interpretable values (i.e. modules that are not too large or too small to interpret biologically). Typically, scale-free fit increases asymptotically with power, so that any power around the asymptote can be chosen based on the connectivity spline. However, in samples in which groups exist (like our high and low drinkers), aberrant distribution of scale-free fit values over power values are often seen. In these cases, a standard power of 6 is recommended by the creators of the WGCNA package.

Spearman correlations were used to determine the relationship between each modules first principal component ("Eigen Gene") and *Last Week Mean Total Ethanol Consumption*. A corresponding p-value < 0.05 was considered to be significant. Module Membership (or correlation with the module Eigen gene) and Intramodular Connectivity (a metric representing the number of connections a single gene has within its network) were calculated for each gene, using functions provided by the WGCNA package (38). Genes with Module Membership > 0.8 and Intramodular Connectivity > 0.9 were considered to be hub genes. Modules with significant drinking level correlations are thought to either drive drinking differences, or to be moderated by alcohol exposure. In either case, they represent important functional frameworks relevant to cortical pathways that mediate drinking behaviors. These modules were also tested for overrepresentation with genes that were significantly differentially expressed between high drinkers and controls, via Fisher's exact test, with a $q_{FDR} < 0.1$ representing significance. Finally, we tested all modules for overrepresentation with genes containing significant SNPs in the GSCAN GWAS on average alcoholic drinks per week (61). For this test, empirical p-values were calculated via permutation, as different gene sets (i.e. human genes vs. mouse genes) were used for each of the original analyses. For each module, random modules of the same size were permuted 10,000 times, and the number of genes containing significant ($p < 5*10^{-8}$) was plotted to create the null distribution. Modules that contained a number of significant SNP-containing genes that fell above the 5th percentile of this distribution, representing a p-value < 0.05, were considered to be significantly overrepresented. Because these p-values were empirical, no multiple testing correction was performed. The GSCAN sample is a meta-analytical sample composed of

941,280 participants from several different samples including UK Biobank, dbGAP, ARIC, MESA, eMERGE, Stroke, BEAGESS, and Jackson Heart Study (https://genome.psych.umn.edu/index.php/GSCAN_dbGaP). All samples were imputed to the Haplotype Reference Consortium (275), using Minimac3 (276) or IMPUTE2 (126). Meta-analysis was performed using rarGWAMA (https://github.com/dajiangliu/rareGWAMA/), on summary statistics obtained from RVTESTS (277) for effect of variants on average number of alcoholic drinks consumed per week (61). Relatedness for family samples, cryptic relatedness for population samples, and population stratification were handled by including genomic principal components or kinship matrices as covariates during the original generation of summary statistics (61). Significance threshold for meta-analysis was $5*10^{-8}$ for variants with Minor Allele Frequency ≥ 1%, and $5*10^{-9}$ for those with MAF < 1% and > 0.1%. Gene-wise p-values were calculated using PASCAL (278).

Expression Quantitative Trait Locus Mapping

According to methods described in Chapter 4, we mapped expression quantitative trait loci (eQTL), or loci at which alleles affect expression values of a given gene, for top candidate genes for ethanol consumption and preference (for ethanol over water and 30% ethanol over 15% ethanol) from *Chapter 4*, using the R *qtl2* package (165). To identify loci that affect the overall expression of such modules, we also performed eQTL analysis on each WGCNA network's Eigen Genes, to identify loci where alleles result in module-wide expression differences. *Flow Cell* and *Cohort* were included as categorical covariates, and *Kinship* was included as a fixed effect, in a linear mixed model. *Kinship* was estimated based on haplotype probabilities (i.e. the probability that each region of DNA was inherited from

each of the 8 progenitor strains, identity by descent), using the Leave One Chromosome Out method (156, 165). Expression was quantified as log-transformed TPM, and genotypes were represented by haplotype probabilities (as they were in the kinship estimation).

Because of the computational intensiveness of the permutation analysis required for determining significance thresholds for each individual analysis (i.e. each gene whose expression is being analyzed) and the similarity of thresholds between traits seen in *Chapter 4*, all loci with LOD peaks > 6 were considered to be suggestive, and those with LOD peaks > 7.58 (the average threshold LOD peak corresponding to $p < 0.05$ in the bQTL-based permutations) were considered to be significant. The 95% Bayesian support intervals for suggestive and significant loci will be compared to those of behavioral QTL mapped for drinking behaviors in these mice in a previous study (*Chapter 4*) to test for overlap between 95% Bayesian support intervals. Overlap between bQTL and eQTL provides evidence that the bQTL effects drinking behavior via a mechanistic pathway related to the gene that the overlapping eQTL regulates.

Functional Overrepresentation

Modules that were significantly correlated with *Last Week Mean Total Ethanol Consumption*, and were significantly overrepresented with GSCAN GWAS significant genes for average drinks per week, or with significantly differentially expressed genes between high drinking DO mice and controls, were tested for functional overrepresentation. All modules with Eigen Gene eQTL 95% Bayesian support intervals that overlapped with bQTL identified for drinking behaviors in these mice were tested for functional overrepresentation, as well. ToppGene's ToppFun function was used to perform gene ontology-based functional overrepresentation testing (279). All overrepresentations with $p < 0.01$ were considered

significant, but those with $q_{FDR} < 0.1$ were considered to be highly significant functions. For larger modules with > 20 highly significant overrepresentation groups per ontological category, the top 20 groups for each category are shown in results tables. All significant results are reported in supplementary tables. Categories were limited to those that included anywhere between 3 and 1,000 genes from a given module. The following categories were tested for overrepresentation: molecular function, biological process, cellular component, human phenotype, mouse phenotype, pathway, and disease.

Results

Differential Expression and Drinking Associations for Individual Genes

Total ethanol consumption and preference values for the 100 highest and 100 lowest drinkers can be seen in Figure 6. (See full distribution of consumption values for all mice in *Chapter 4*.) One high drinker was removed due to potential sample mix up before sequencing, however all other samples were retained after quality control. These groups varied across experimental cohorts, but were not unbiased, and expression values appeared to group with *Cohort* in principal component plots (Figure 3 & 4c). As described in the *Methods* section, we either corrected for *Cohort* or included it as a covariate in all of our analyses. Furthermore, due to a breach in protocol, we discovered that flow cell on which the samples were sequenced corresponded with drinking groups, such that the highest drinkers, controls, and 28 of the highest drinking low drinkers were on the first flow cell, and the remaining 72 lowest drinkers were on

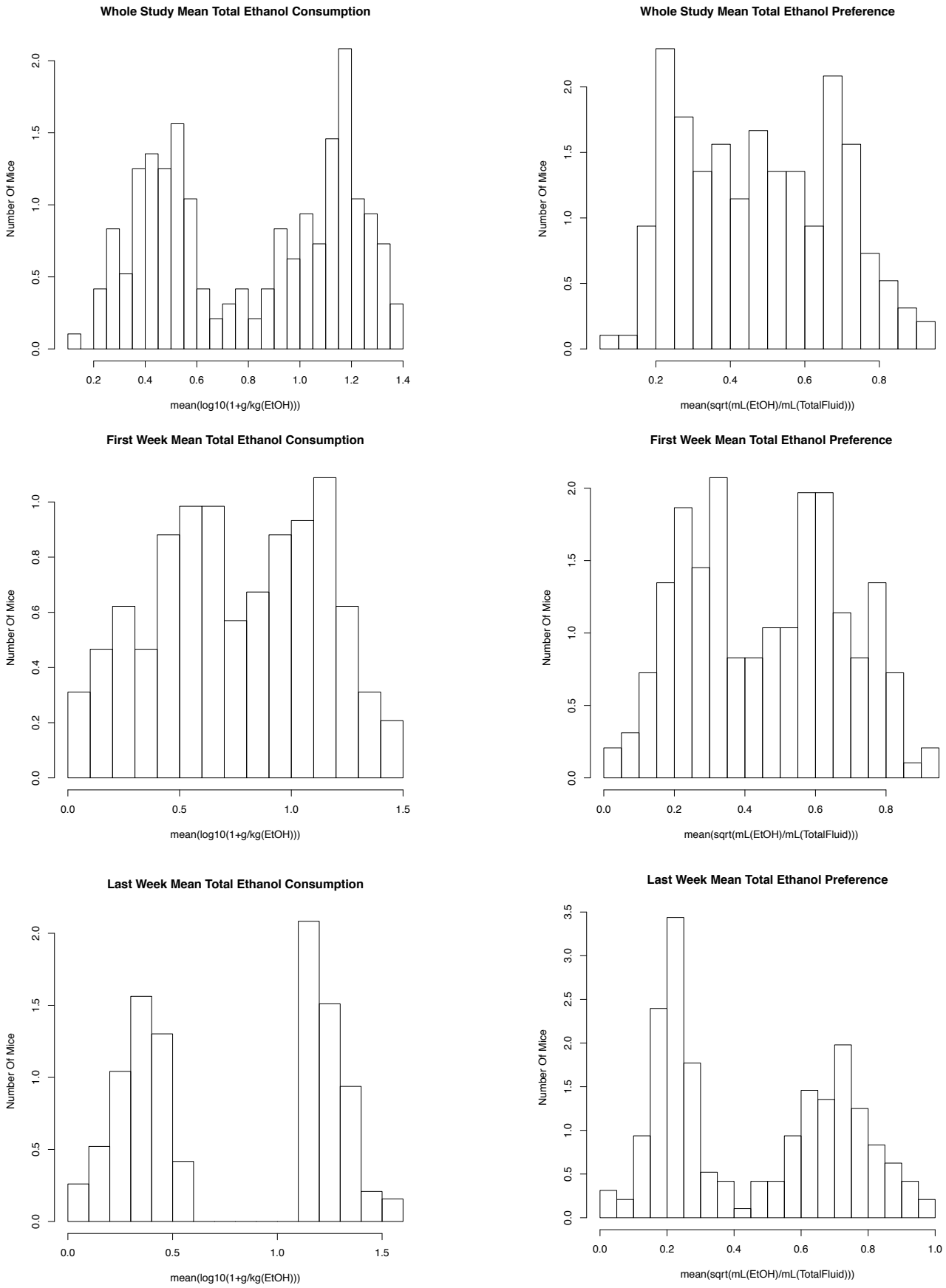


Figure 6: Histograms of total ethanol consumption (left) and preference (right) values, across each tested time interval, indicated at the top of each plot. Distributions appear especially bimodal for the last week time interval, because of the sampling from extremes, based on last week mean total ethanol consumption values, for sequencing.

the second flow cell. This was also evident in the principal component plots of expression values (Figure 4c). Although correcting and controlling for flow cell would likely reduce our power, we decided to do so, in order to avoid the capture of false positive results due to technical variation. After using Limma to correct for Flow Cell and Cohort, these clusters disappeared in principal component plots, indicating adequate correction for the correct covariates (Figure 4c). However, it was not necessary to control for flow cell differences for DEseq comparisons between high drinkers and controls, as these two groups were run on the same flow cell.

All results for gene expression level covariation with ethanol drinking behaviors, and for differential expression between high drinkers and controls for all genes can be seen in Table S1. Of the 17,709 genes analyzed, 376 analyses failed to converge in regressions against *Last Week Mean Total Ethanol Consumption*. Therefore, when comparing DEseq results to *Last Week Mean Total Ethanol Consumption* correlation results, we only consider those that successfully yielded results in both analyses (N_{genes}=17,333). For the differential expression analysis between high drinkers and controls, 1,171 genes were significant; whereas 6,968 genes significantly predicted *Last Week Mean Total Ethanol Consumption* across all 200 drinkers. Of these, 564 genes displayed significance in both analyses. After removing genes that had missing values in DEseq due to failure of that gene to meet the normalization algorithm assumptions (N_{failed}=2,038), in order to compare differences in results between analyses, 6,371 of the significantly ethanol consumption-associated genes remained. This means that 607 genes were differentially expressed between high drinkers and controls, but did not significantly predict ethanol consumption, and the opposite was true for 5,807 genes. To determine functional trends, gene ontology overrepresentation analysis

was run on each of three gene sets, using ToppGene (279): genes that were both significantly differentially expressed and significantly predictive of ethanol consumption levels; genes that were only significantly differentially expressed; and genes that were only significantly predictive of ethanol consumption levels. The same parameters were used for this overrepresentation as those used for the ToppGene analysis of WGCNA modules. However, because this analysis is focused on finding general functional trends in groups of genes as opposed to attributing function and pathology to a specific network, only the categories of molecular function, cellular component, biological process, and pathway were searched. All results with $p < 0.01$ were considered significant (Table S4), and the top 20 highly significant results ($q_{FDR} < 0.1$) for each ontological category can be seen in Table 2a-c.

Gene Network Analysis

WGNCA results yielded a total of 33 modules, and one "module" consisting of 7,709 genes that were not assigned a module, meaning they were removed from modules during clustering, due to low module membership (correlation with the Eigen gene) (Figure 7). Module Membership, Intramodular Connectivity, and Hub Gene status for each gene are represented in Table S1. Results for each module's association with ethanol drinking behaviors, and overrepresentation for differentially expressed genes between high drinkers and controls, are presented in Table S3. Seventeen modules had Eigen Genes significantly correlated with at least one of Total Ethanol Consumption or Preference for the first week of IEA, last week of IEA, or the whole study (Figure 9). Not surprisingly, given that gene expression data was collected after the last week of drinking, none of the modules were significantly correlated with *First Week Mean* phenotypes. These 17 modules were carried

forward for testing for overrepresentation of differentially expressed genes between high drinkers and controls (from the DEseq2 analysis) and significantly drinking level-associated genes from GSCAN, and for overlap between Eigen Gene eQTL and bQTL for ethanol drinking phenotypes identified in the previous chapter. Seven of these modules (Light Yellow, Brown, Turquoise, Magenta, Yellow, Pink, and Black) were significantly overrepresented with differentially expressed genes, and two modules (Turquoise and Blue) were significantly overrepresented with genes harboring GSCAN GWAS significant SNPs (Table 3a,c).

Finally, eight significant eQTL were identified for module Eigen Genes, each belonging to a unique module, and 50 suggestive QTL were identified (Figure 8, Table 4). Five of these eQTL overlapped with support intervals of behavioral QTL identified for Last Week Mean Total Ethanol Consumption, Last Week Mean Total Ethanol Preference, and First Week Mean Total Ethanol Consumption. The eQTL for six modules' Eigen Genes overlapped with behavioral QTL identified in the Chapter 4. Modules Blue, Light Green, and Pale Turquoise overlap with the Chr1 bQTL for *Last Week and Whole Study Mean Total Ethanol Preference*; modules Brown and Light Green overlap with the Chr6 bQTL for *First Week Mean Total Ethanol Consumption*; and module Dark Turquoise overlaps with the Chr16 bQTL for *Last Week Mean Total Ethanol Preference*. However, the Bayesian support intervals for five of these eQTL (those on Chr1 and Chr6) are extremely large (>130Mbp), so these results are to be interpreted with caution. In accordance with the criteria discussed in the *Methods* section of this chapter, six modules were chosen to be carried over for functional overrepresentation analysis: Brown, Turquoise, Light Green, Blue, Dark Turquoise, and Pale Turquoise. Complete tables of overrepresentation results with $p < 0.01$ can be seen in Table

S4. All highly significant results can be seen in Table 2. Every module was significantly overrepresented within each functional category. Although, not all modules had highly significant overrepresentations in every category, every module was highly significantly overrepresented in at least 2 categories (as is evident in Table 3), indicating that our analysis was successful at identifying groups of ethanol-associated and functionally related genes.

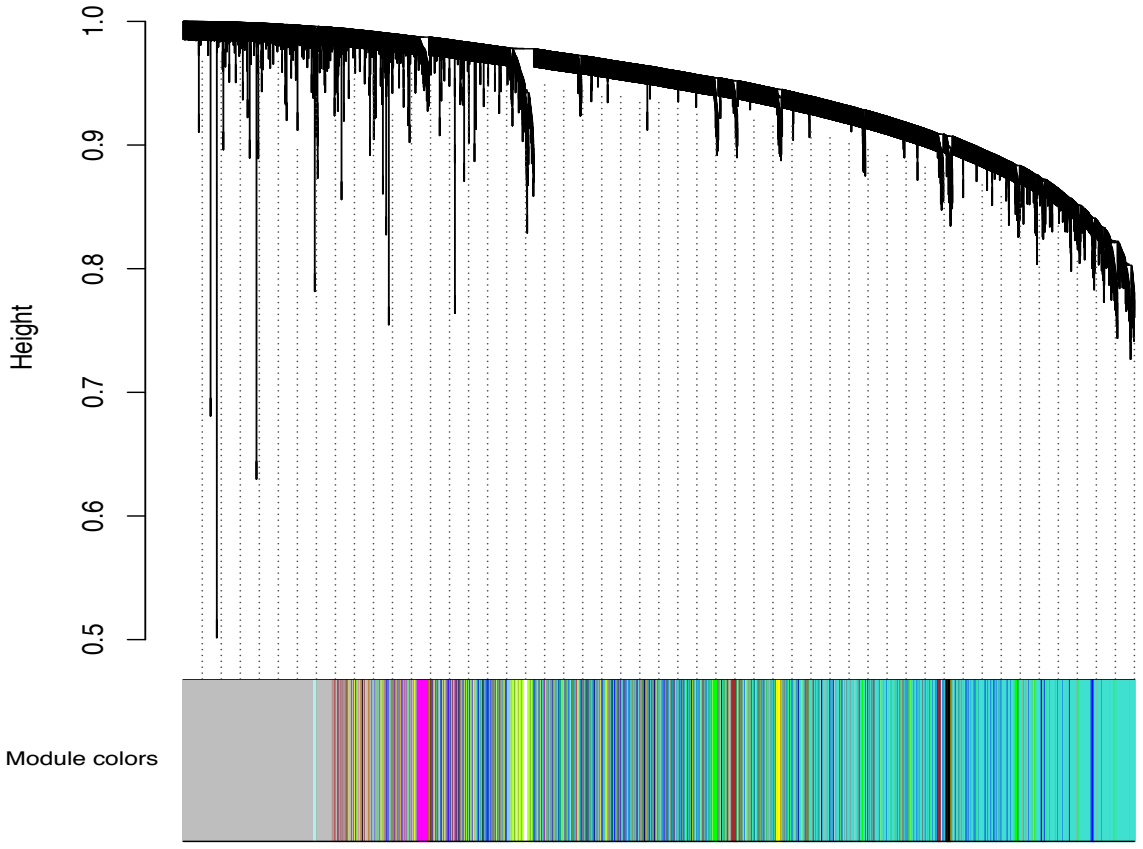
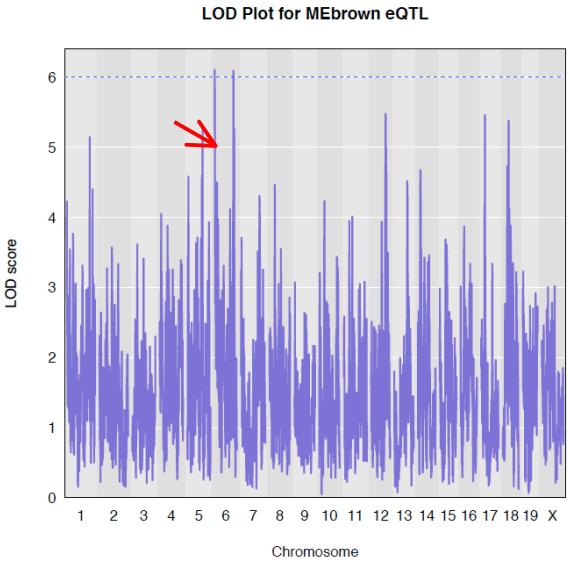


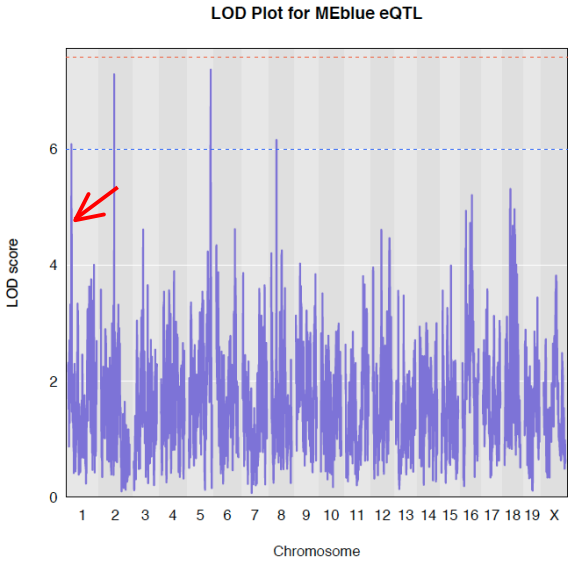
Figure 7: Cluster dendrogram representing module relationships. Each leaf of the tree represents a single module, while branching represents correlation. Height represents distance (or the inverse of correlation). Based on an algorithm that chooses cut heights across the tree, based on height differences between branches at each point across the tree, branches are cut to form modules, indicated by colors below the tree.

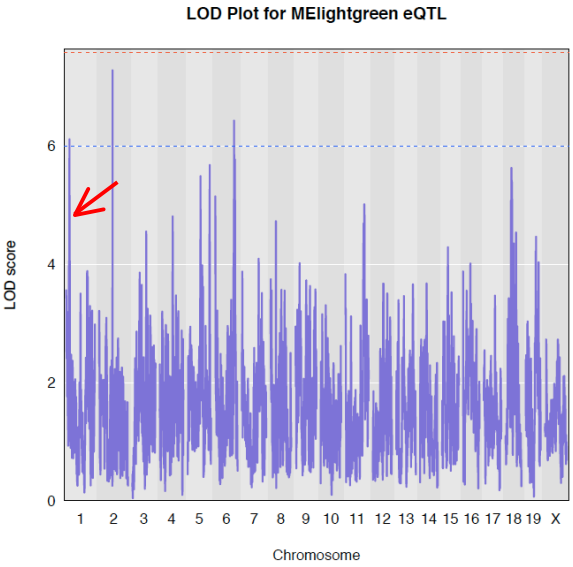
Table 4: Module Eigen Gene eQTL

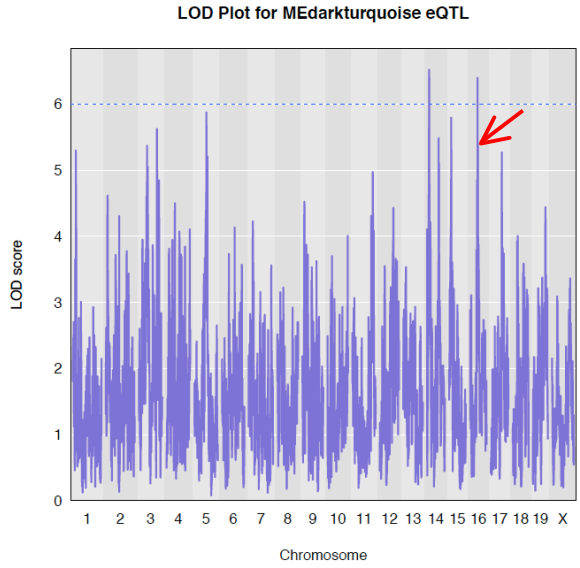
Module	marker	LOD Score	Chr	Position (Mbp)	Downstream 95% Bayesian SI	Upstream 95% Bayesian SI
MElightcyan	UNCHS005711	9.52	2	90.5673	90.5656	90.5673
MEviolet	UNC12205668	8.85	6	138.3612	137.3680	139.2137
MEgreen	UNC9440121	8.66	5	72.2247	67.5711	73.9091
MEorange	c4.loc58.78	8.36	4	135.8440	135.7429	135.9371
MEgrey60	c6.loc63.83	8.33	6	139.1233	135.1719	139.2640
MEdarkred	UNCHS005711	8.31	2	90.5673	10.8933	90.5673
MEdarkolivegreen	UNC16043985	7.62	9	27.3563	26.3948	27.8608
MEpurple	cX.loc34.87	7.62	X	84.9481	75.0489	87.3748
MEtan	c5.loc71.61	7.53	5	146.6990	144.4243	146.8857
MElightyellow	cX.loc33.44	7.38	X	75.8546	74.3731	87.3278
MEblue	c5.loc71.6	7.37	5	146.6973	144.4352	146.8572
MEdarkolivegreen	UNC25242206	7.31	15	27.1487	13.4260	28.4144
MEblue	UNCHS005711	7.29	2	90.5673	3.1767	166.5001
MElightgreen	UNCHS005711	7.28	2	90.5673	3.1767	179.7984
MEdarkgreen	UNCHS005711	7.24	2	90.5673	3.1767	168.5210
MElightyellow	UNCHS016304	7.21	5	147.2071	139.0748	147.2197
MEmidnightblue	UNC9776212	7.14	5	100.4211	100.2420	102.8197
MEblack	UNC29114578	7.12	18	41.6204	38.9586	42.4633
MEroyalblue	UNC16011855	7.08	9	24.6765	23.3848	27.3563
MEred	c5.loc71.61	6.94	5	146.6990	144.4016	146.8572
MEsaddlebrown	UNCHS015190	6.87	5	101.7593	100.2420	102.9887
MEsteelblue	UNC9816473	6.82	5	102.9887	99.5884	147.6169
MEskyblue	UNCHS014691	6.81	5	71.9714	67.6002	139.1954
MEpurple	JAX00486143	6.81	2	28.3310	26.2775	77.1019
MEwhite	c6.loc61.18	6.76	6	135.9998	133.9692	137.7865
MEpaleturquoise	UNC22338724	6.71	13	29.4673	27.8470	30.6940
MEcyan	UNCHS005711	6.61	2	90.5673	3.1767	180.1845
MEblack	c4.loc58.76	6.55	4	135.8216	135.7295	135.9371
MEdarkred	UNC16043985	6.54	9	27.3563	23.3848	40.7505
MEdarkturquoise	UNCHS037737	6.52	14	27.8875	25.5451	91.6242
MEgrey	JAX00147136	6.50	6	135.2173	3.2002	139.4509
MEviolet	UNCHS032383	6.47	11	114.6716	106.6006	115.1819
MEdarkorange	UNC23726899	6.47	14	25.6757	25.5613	88.0969
MEgreenyellow	JAX00345858	6.44	12	110.1658	97.5934	112.2763
MEgreenyellow	UNCHS005711	6.43	2	90.5673	3.1767	172.8370
MElightgreen	c6.loc60.53	6.43	6	135.0735	3.2002	139.4509
MEdarkred	c5.loc71.61	6.43	5	146.6990	144.4016	146.8448
MEpurple	c12.loc0.8	6.41	12	6.0204	3.9307	70.5633
MEdarkorange	UNC25242206	6.40	15	27.1487	14.3397	30.1815
MEdarkgreen	UNCHS012241	6.40	4	104.0428	101.1600	104.3381
MEdarkturquoise	UNC26725019	6.39	16	44.2626	39.9709	48.5580
MEgreenyellow	UNC10385175	6.39	5	146.7027	24.2063	147.8497
MEblack	JAX00171884	6.39	9	60.7010	60.1659	61.3391
MElightcyan	UNC9799387	6.28	5	102.0051	99.3871	147.0045
MEorange	UNCHS019510	6.27	7	17.0169	3.1577	95.6569
MEpaleturquoise	UNC386248	6.25	1	32.1956	10.1715	160.1762
MEsaddlebrown	UNCHS005711	6.24	2	90.5673	3.4337	180.1137
MEred	UNCHS005711	6.19	2	90.5673	3.1767	161.1323
MEdarkorange	UNC16043985	6.19	9	27.3563	22.8308	41.1797
MEblue	UNC14510156	6.16	8	36.3829	3.1232	102.7612
MEroyalblue	UNC29056323	6.15	18	37.1504	35.7468	38.2154
MEmagenta	UNC23794297	6.13	14	31.5658	30.7098	33.1612
MElightgreen	JAX00246131	6.12	1	33.7401	30.4238	164.2401
MEbrown	UNCHS016395	6.10	6	3.4501	3.2002	138.3612
MEblue	JAX00246131	6.09	1	33.7401	24.5680	187.6188
MEdarkorange	UNC385804	6.03	1	32.1123	31.3310	34.4871
MEblack	UNCHS029800	6.01	11	10.4093	9.5554	16.1413
MEgreenyellow	UNCHS011115	6.00	4	44.4351	20.1889	155.3592

Significant (LOD > 7.58), and suggestive (LOD > 6) module Eigen Gene eQTL with the name and chromosomal position in Mega base pairs (Mbp) of the top-scoring marker for each locus. 95% Bayesian support interval calculations were used to determine the upper and lower bounds (in Mbp) of each QTL. Module names are highlighted in grey to indicate that those eQTLs overlapped with behavioral QTLs for ethanol drinking behaviors.









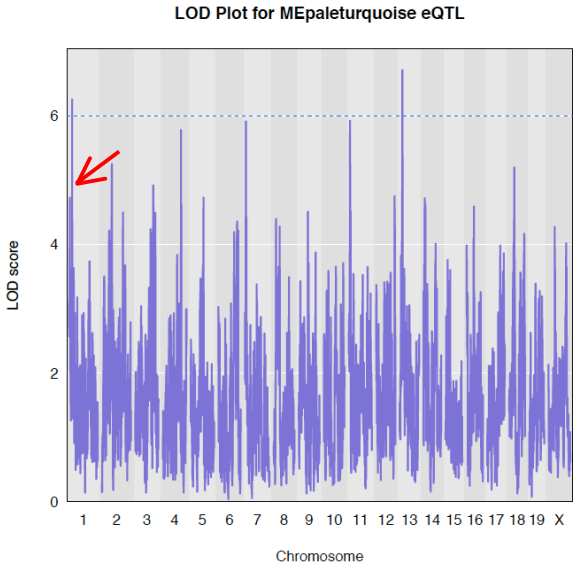
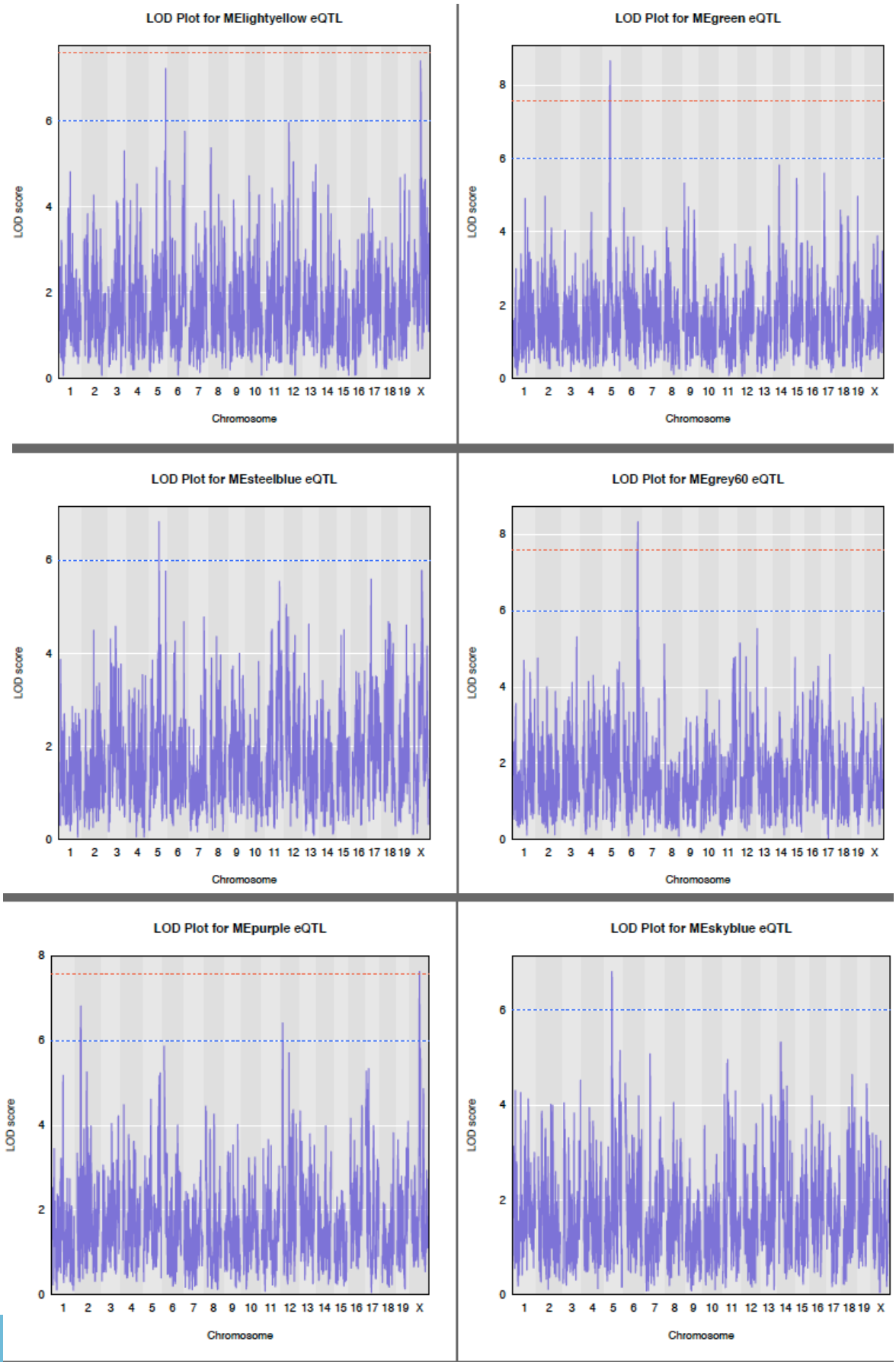
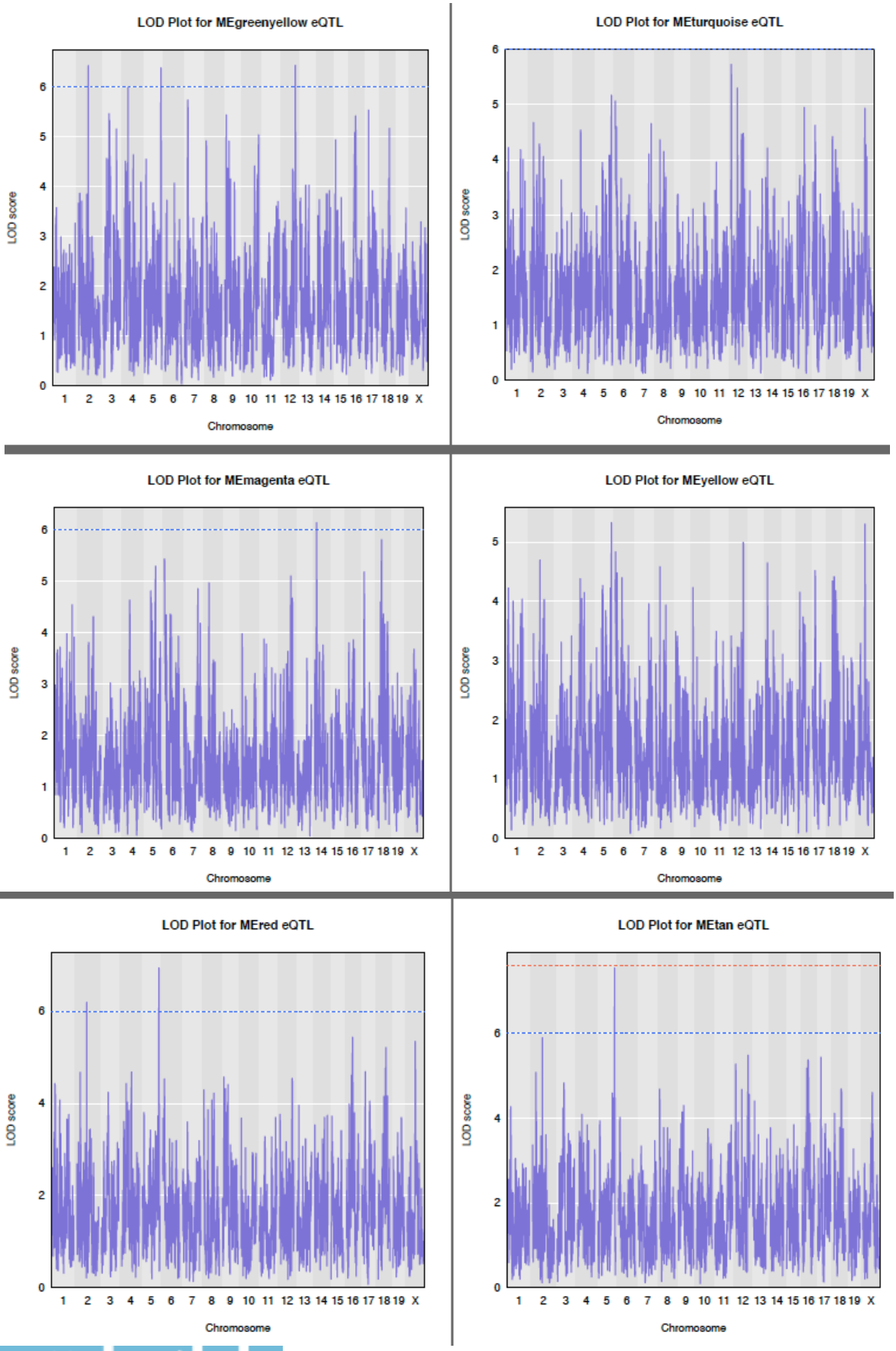
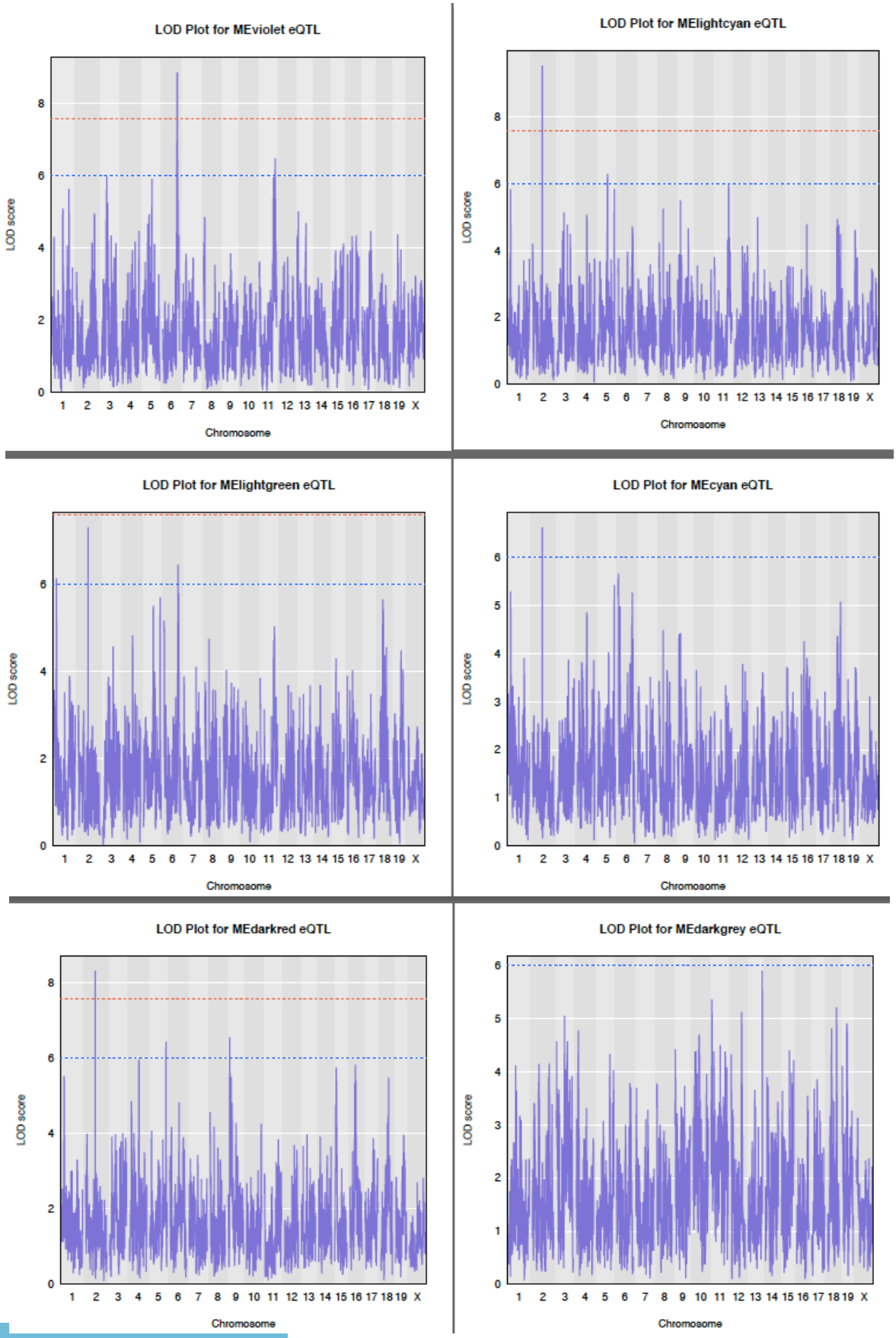
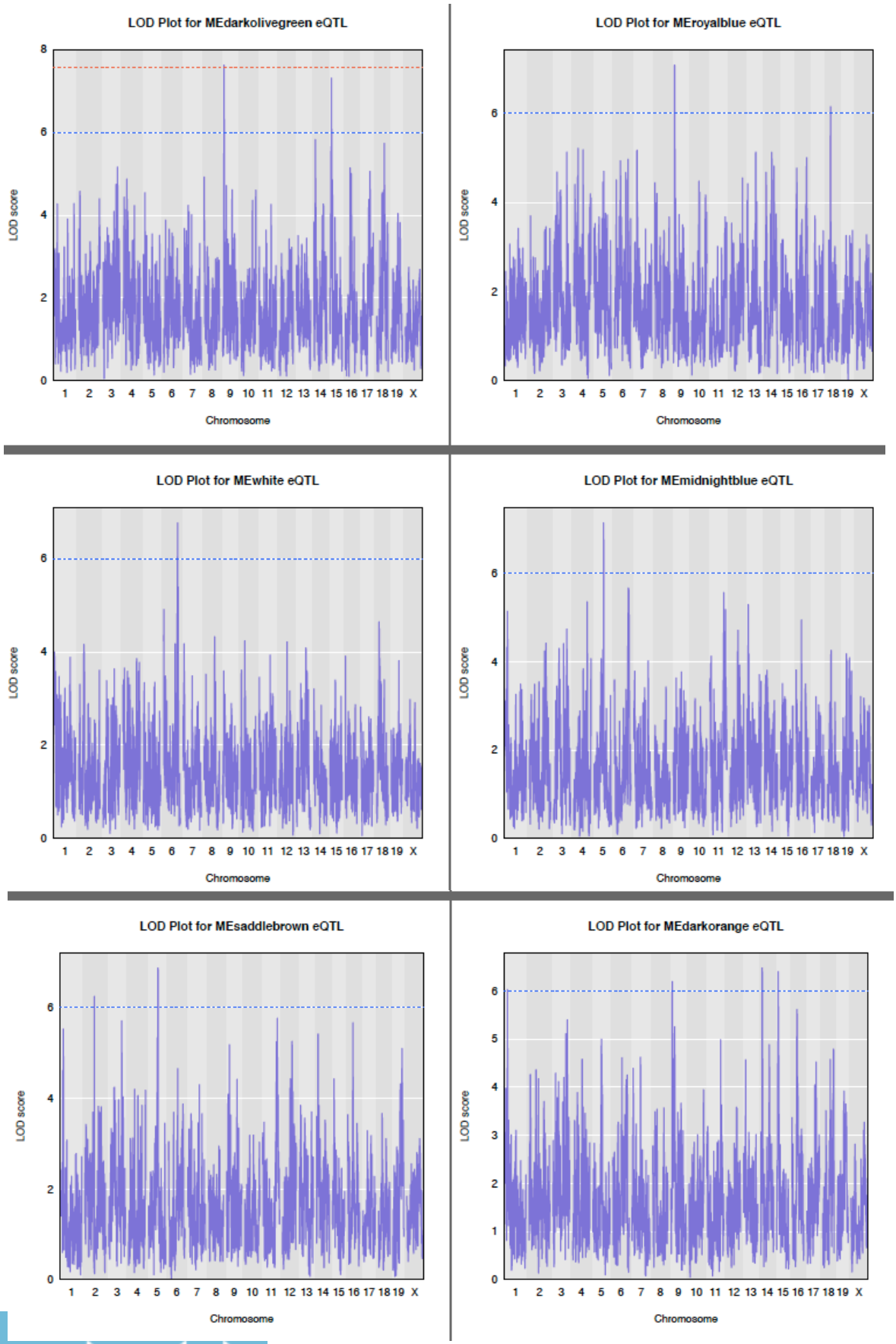


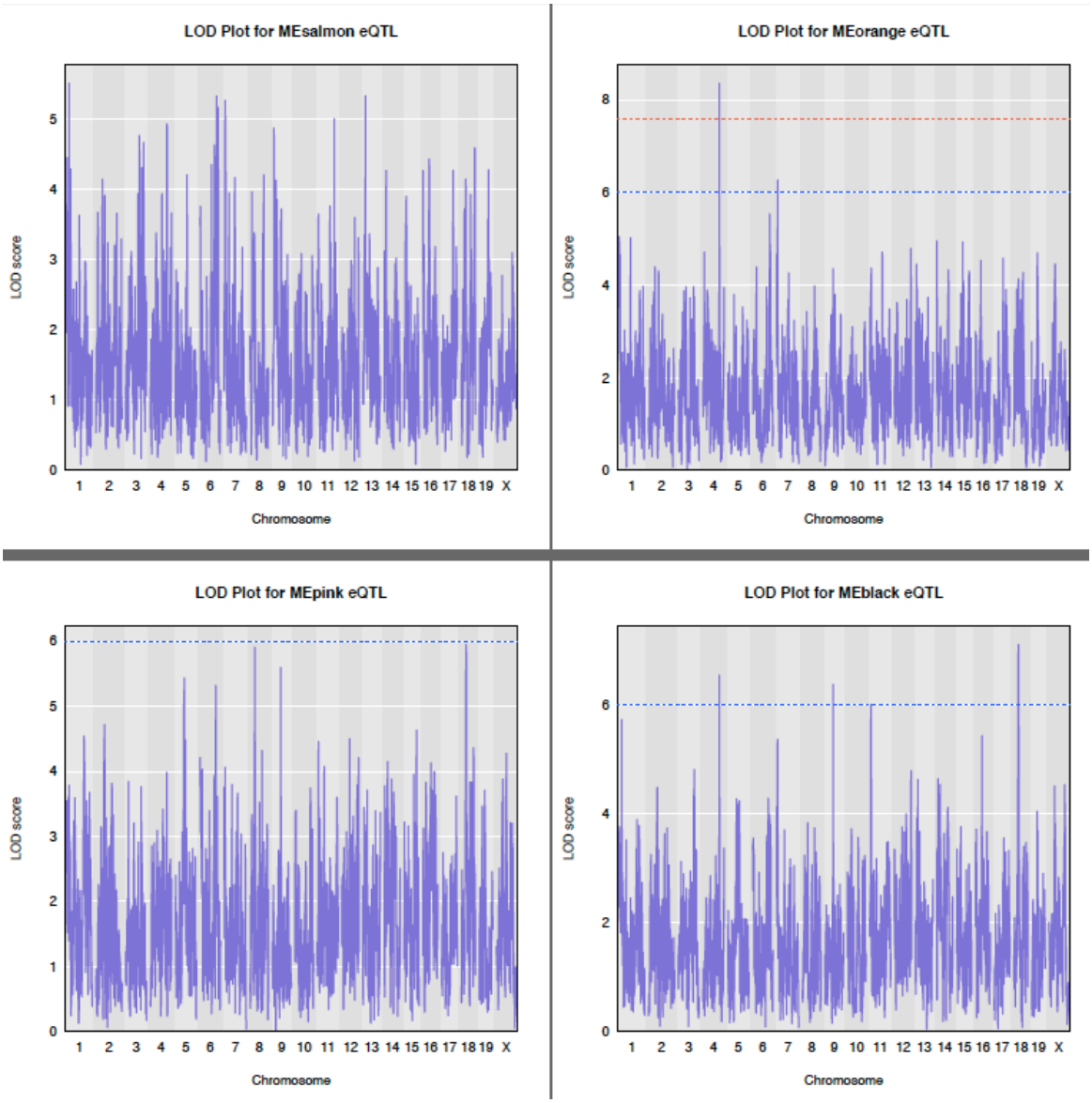
Figure 8: LOD scores across all chromosomes yielded by module Eigen Gene eQTL analysis, for each WGCNA module. The first 5 LOD plots (shown on this page) are for modules that had an eQTL that overlapped with a behavioral QTL for an ethanol drinking behavior. (Plots for remaining modules are below) These specific eQTLs are denoted by red arrows. Blue and red dotted lines represent suggestive and significant LOD score thresholds, respectively. Scores that fall above those lines are either suggestive or significant, respectively.











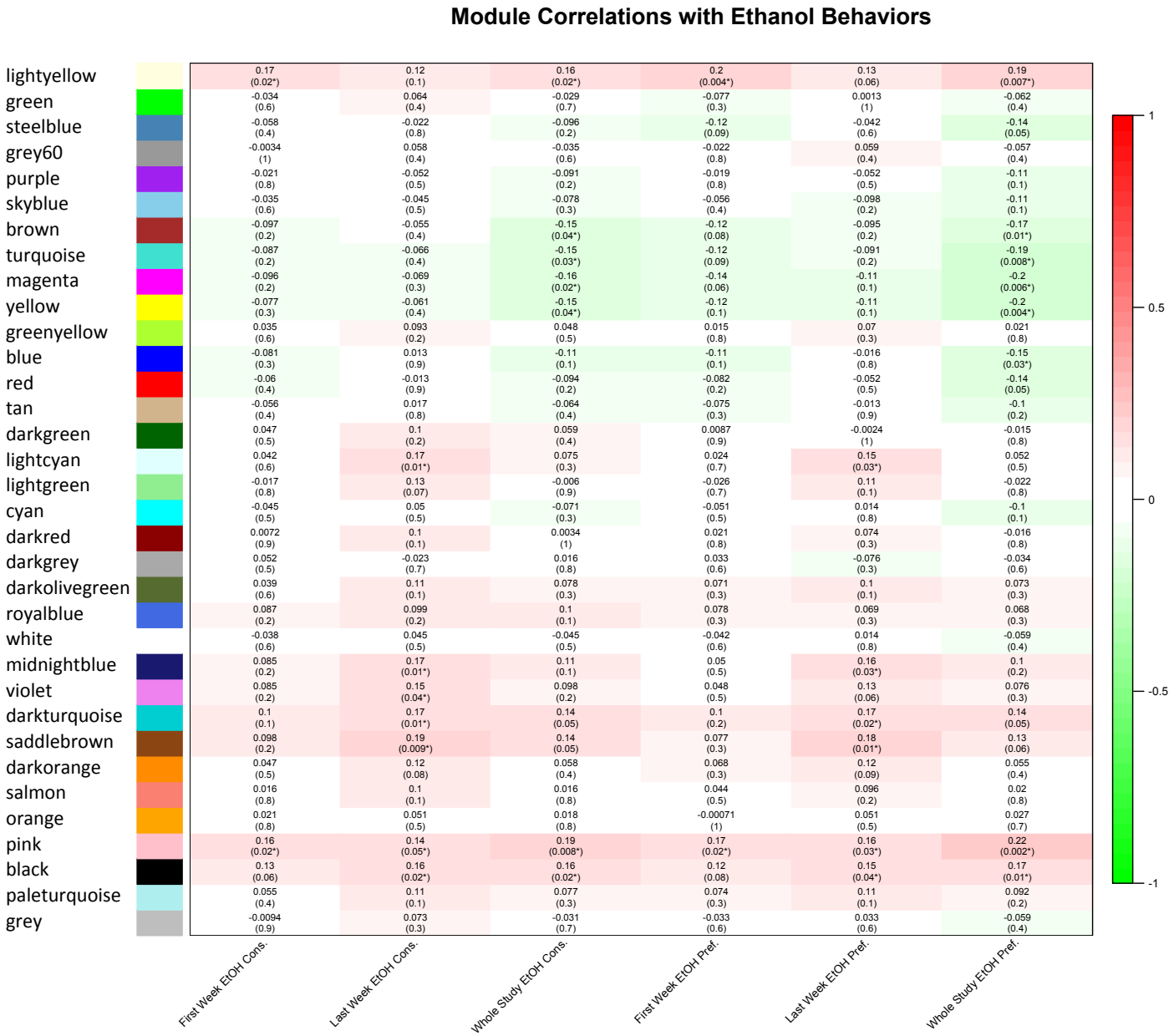


Figure 8: Module Eigen-Gene correlation values with each drinking behavior tested. Each row represents a module, indicated by color on the left-hand side of the plot. Green cell coloring indicates negative correlations, and red coloring indicates positive correlation, with darker colors representing stronger correlations. Values within cells represent the correlation coefficient above the p-value for that correlation. Asterisks are placed next to significant p-values (< 0.05).

Discussion

Differential Expression and Ethanol Consumption Association

Of the 1,171 genes that were significantly differentially expressed between controls and high drinkers, 607 did not predict drinking levels amongst the high and low drinkers, whereas 5,807 of the 6,371 genes that predicted drinking levels were not significantly differentially expressed. We anticipated seeing genes fitting three of these descriptions. Genes fitting the latter description are likely ethanol-regulating (meaning their expression levels moderate drinking behaviors), as opposed to being ethanol-regulated (meaning ethanol exposure moderates expression levels). In contrast, genes that were significantly differentially expressed between high drinkers and controls but did not predict ethanol consumption may be highly sensitive to ethanol, creating a ceiling effect, rendering different expression levels between ethanol-naïve and ethanol-exposed mice, and similar expression levels across mice that consumed low and high amounts of ethanol. Therefore, these genes would still be classified as ethanol-regulated, but with extreme ethanol sensitivity or tight regulation by feedback loops, creating a ceiling effect. Top results from functional analysis revealed that these genes are highly overrepresented with functions related to: transmembrane cation (specifically potassium) transport; localization in all parts of neurons (axons, soma, and dendrites); calcium signaling; alpha adrenergic signaling; insulin secretion; cardiac conduction; and muscle contraction (Table 2a). Neuronal signaling, ion transporters, signaling cascades are just the types of pathways that one would be expect to be affected by alcohol exposure.

For the 607 genes with more ethanol dose-dependent expression levels, or those that were significantly differentially expressed and predictive of consumption levels, several of the top functional enrichment categories were particularly involved in neuronal function. These categories included: transmembrane transporter activity; neuron projection development; and localization in the myelin sheath. There were also several categories involving cellular metabolism, and energy production and release, including: mitochondrial activity; cellular respiration; citrate and tricarboxylic acid metabolism; oxidation; pyruvate dehydrogenase activity; glucose metabolism; and ATP synthesis. Interestingly, these genes are also involved in processes related to Parkinson's, Alzheimer's, and Huntington's diseases, all of which involve neuron death and myelin degeneration (Table 2b). It could be that one component of alcohol's toxicity to neurons is the perturbation of highly sensitive, energy-related pathways that could be related to neuronal and myelin integrity, or to neuron projection maintenance and development. These findings are supported by prior research that has shown significant death, demyelination, and synaptic remodeling of neurons, and ethanol-regulated gene expression changes, in the Prefrontal Cortex of long-term alcoholics (280) (30) (96). They also indicate that these pathways could even be disrupted by long-term relatively low levels of chronic alcohol consumption.

Finally, the 5,807 genes that we believe to be ethanol-regulating, due to their significant association with ethanol consumption levels but lack of significant differences between high drinkers and controls, top functional enrichment displayed many protein-level regulatory pathways (Table 2c). Specifically, these functions are centered around protein-level and activity control: RNA processing, ubiquitination, proline hydroxylation, GDP binding and GTPase activity, and proteolysis. We hypothesize that these associations may be

stem from robustness against ethanol-induced perturbation of protein levels and activity, which may be protective against increases in consumption and the development of dependence. The remaining top functional categories were largely related to: hydrolase activity; membrane trafficking and vesicle-mediated transport; and immune function. Membrane trafficking and vesicle-mediated transport are particularly interesting categories, as these functions are involved in intracellular signaling. Just as with the protein-processing pathways, there may be a certain robustness to these pathways against ethanol-induced change. This quality could render an individual less prone to tolerance, which is mediated by changes in transmembrane receptor density, and craving, which is mediated by synaptic plasticity in response to ethanol-induced changes to normal signaling (249) (8). Finally, immune function has been found to be compromised by long-term alcohol consumption, and neuroinflammation and immune-related changes have been found to be increased in alcoholics (265) (264) (28) (106).

It is important to note that some of the differences in results between DEseq2 and the correlations between gene expression values and *Last Week Mean total Ethanol Consumption* and *Preference* may differ due to lack of power in DEseq2 compared to the phenotype regressions, as there were only 20 samples in the control group. This could have resulted in fewer significant genes in DEseq, making it appear as though many genes were ethanol-regulating (i.e. significantly predicted consumption and preference values) but not ethanol-regulated (i.e. did not differ between controls and high drinkers). However, the algorithms used DEseq2 were specifically designed to be robust against both small samples sizes and difference in sample sizes between groups (273). Therefore, we believe that the differences between these two analyses are biologically relevant.

Gene Expression Networks

Our analysis revealed 17 gene expression networks in the PFC whose Eigen Genes were significantly correlated with *Last Week Mean Total Ethanol Consumption*. This indicates that the overall expression levels of the genes in those modules are either ethanol-regulated or ethanol-regulating. Seven of these modules were also overrepresented with genes that were differentially expressed between high drinkers and controls, indicating that these modules may be ethanol-regulated. One of these seven modules (Turquoise) and four additional modules were significantly overrepresented with genes displaying significant associations with "average drinks per week" in the GSCAN GWAS. These results suggest that Turquoise is an ethanol-regulated module that contains genes that also regulate alcohol consumption. We believe that the sensitivity of these genes to ethanol likely mediates the relationship between the associated SNPs and drinking levels in humans. Finally, five modules' Eigen Gene eQTLs overlapped with bQTLs identified for ethanol drinking behaviors in these mice in the previous chapter, two of which (Blue and Light Green) were overrepresented for GSCAN GWAS significant genes, and one of which (Brown) was also significantly overrepresented with differentially expressed genes between high drinking and control DO mice. Because Blue and Light Green show evidence for mediating the effects of genetic variants on drinking, but are not significantly overrepresented with differentially expressed genes between highly drinking and non-drinking mice, these results suggest that affective variants moderate basal expression levels or functionality which then affect drinking behaviors. Similar to Turquoise, we theorize that the sensitivity of genes in the Brown module to ethanol may mediate the effects of the variants in the Chr1 bQTL interval

on long-term ethanol consumption. However, as expressed in the Methods, the eQTL intervals for four of these modules (Blue, Light Green, Pale Turquoise, and Dark Turquoise) were quite large (> 130Mbp), so results should be interpreted with caution.

The Turquoise module was highly significantly overrepresented for pathways largely related to the following functions: cognitive impairment and developmental delays; microcephaly and abnormal brain morphology; muscular hypertonia and atrophy; abnormal facial feature morphology; abnormal eye movement; ubiquitin and ligase activity; ATP- and GTP-binding; protein modification; organelle and chromosomal organization; and Golgi vesicle transport (Table 3a). This module's hub genes were *Appbp2* (amyloid beta precursor protein (cytoplasmic tail) binding protein 2), *Pafah1b1* (platelet-activating factor acetylhydrolase, isoform 1b, subunit 1), *Dnm1l* (dynamin 1-like), and *Mtpn* (myotrophin). Expression levels of all of these genes are significantly associated with *Last Week Mean Total Ethanol Consumption*, but surprisingly (given this module's overrepresentation with differentially expressed genes), only *Pafah1b1* is differentially expressed between high drinking mice and controls (Table S1). One possible explanation for this is that basal levels of expression of the other three hub genes affect the ethanol-sensitivity of the remaining genes in the module, one of which is *Pafah1b1*. All four hub genes were present in highly significantly overrepresented functional groups for this module, and even shared some functional categories. *Appbp2* and *Dnm1l* are both involved in phosphatase and hydrolase activity, and are known to be located in vesicles. Furthermore, prior literature has shown support for the association of *Dnm1l* and Beta-Amyloid (the precursor of which binds to *Appbp2*, hence this gene's name) with neurodegenerative diseases (281, 282). *Dnm1l* and *Mtpn* are both associated with protein complex assembly, and *Mtpn* and *Pafah1b1* are both

localized in axons. Relevant to both of these functional categories, *Mtpn* moderates the dimerization of the protein NFκB's (nuclear factor kappa B) subunits. This protein is primarily known for activating the transcription of cytokines as a function of immune response, but it is also associated with drug-related reward, response to stress, and Beta-Amyloid levels in the brain (283, 284). Additionally, both NFκB and Dnm1l are upregulated by cocaine exposure in microglia (285, 286). Finally, *Dnm1l* and *Pafah1b1* share several functions, including: cellular response to stress; muscular hypotonia; organelle fission; cell part morphogenesis; microcephaly; brain atrophy (i.e. neurodegeneration); and seizures. It appears that the overarching theme of functionality for this network is neurodegeneration. In fact, a neuropathological relationship between Alzheimer's and Alcohol Dependence has already been suggested by Venkataraman et al. (287). The proposed shared mechanisms are neuroinflammatory in nature and involve Beta-Amyloid and TLR4 (toll-like receptor 4), which is responsible for promoting the NFκB signaling pathway in microglia (287).

The Brown module was highly significantly overrepresented with functional categories that largely represented DNA damage repair, apoptosis (particularly in response to DNA damage), as well as cognitive impairment and developmental delays (Table 3b). Of the six hub genes for this module (*Brd8*, *Ino80*, *Fam214a*, *Elac1*, *Osbpl2*, and *Zbtb25*), two of them, *Brd8* (bromodomain 8) and *Ino80* (INO80 complex subunit), were observed in highly significant functional groups. Two of these functional groups contained both of these genes: "transferase complex" and "INO80-type complex" (Table S4). *Ino80* is involved in chromatin remodeling, and consequently the moderation expression levels of associated genes, and *Brd8* is involved in hormone-moderated transcription activation (288, 289). *Ino80* also appears in several functional groups related to DNA repair (Table S4). Neither gene's

expression levels are significantly differentially expressed between high drinking and control mice. However, *Ino8*'s expression levels are significantly correlated with *Last Week Mean Total Ethanol Consumption*. A likely mechanism of association with ethanol consumption is that basal expression level differences in this gene directly moderate expression levels of the other genes in this module (via its role in chromatin remodeling), many of which are differentially expressed between high drinker and controls. The *Ino80*-moderated basal expression levels of these genes may affect the sensitivity of related mechanisms to ethanol exposure. For instance, if a mouse already expresses a gene at low levels, and ethanol decreases the expression of that gene, the corresponding mechanisms may be compromised more so than they would be for a mouse with basally higher expression levels. Functional overrepresentation results suggest that these corresponding mechanisms are involved in pathways related to DNA repair and, like Turquoise, cognitive ability. In fact, human GWASs have revealed associations between *Ino80* and intelligence and cognitive ability (170). This is indicative that cognitive ability and drinking behaviors may therefore involve some of the same pathways. Furthermore, another study has suggested that dopaminergic neuronal plasticity in the midbrain mediates prefrontal cortex-driven cognitive control over task-related behaviors in (8). It is well known that dopaminergic pathways are involved in drinking behaviors (8), so it is possible that these shared pathways between cognition and drinking behaviors are both associated with dopaminergic signaling to the prefrontal cortex.

Like Turquoise, the Blue module was highly significantly overrepresented with functions related to microcephaly, cognitive impairments and developmental delays, abnormal facial feature morphology, and muscular hypertonia, all features of syndromes characterized by developmental delays. Additionally, it was highly significantly

overrepresented pathways associated with: synaptic signaling; axon guidance; vesicular transport; cell-cell signaling (via several types of molecules, including hormones, ions, neurotransmitters, growth factors); mood disorders; actin-related processes; cellular and tissue development and organization; ligase activity; and oxidoreductase activity. This modules' hub genes were: *Nf1* (neurofibromin 1), *Trip12* (thyroid hormone receptor interactor 12), *Gapvd1* (GTPase activating protein and VPS9 domains 1), *Ash1l* (ASH1 like histone lysine methyltransferase), *Cnot1* (CCR4-NOT transcription complex, subunit 1), and *Aff4* (AF4/FMR2 family, member 4). *Nf1*, *Trip12*, and *Cnot1* were all in groups representing: transcription factor activity, cellular response to DNA damage, and macromolecular catabolism. *Nf1*, *Ash1l*, and *Aff4* were in functional groups related to localization on chromosomes; and *Nf1*, *Trip12*, and *Aff4* were related to embryonic development. *Nf1* and Gapvd1 are both involved in the promotion of hydrolysis of GTP by Ras (290, 291). Mutations in this gene are associated with neurofibromatosis, characterized by the formation of neuronal tumors (292). *Cnot1* encodes a subunit of a transcription complex and is involved in the deadenylation of mRNA, transcriptional repression, and mRNA degradation (293). Somewhat similar in functionality, *Ash1l* is a histone methyltransferase, *Aff4* encodes a subunit of a complex (super elongation complex) that catalyzes RNA polymerase II activity, and *Trip12* is involved in the degradation of proteins via ubiquitination, all functions that regulate gene expression and protein levels (294-296). Unexpectedly, none of these genes' expression levels were significantly correlated with ethanol consumption in our mouse sample, and were not differentially expressed between the high drinking mice and controls. Although the expression levels of genes in this module do not appear to be ethanol-regulated or ethanol-regulating, it is overrepresented with genes associated with alcohol consumption

in the GSCAN GWAS, and its Eigen Gene eQTL overlaps with a bQTL on Chr1 for *Last Week* and *Whole Study Mean Total Ethanol Consumption* in our DO mouse sample. Furthermore, several of Blue's hub genes have shown significant associations with substance-related and psychiatric traits in human GWAS. In the GSCAN GWAS dataset, *Nf1* is significantly associated with average drinks per week, and *Gapvd1* is associated with smoking initiation. *Gapvd1* is also associated with smoking status in another GWAS sample (170). Finally, *Cnot1* is significantly associated with Schizophrenia in several GWAS datasets (170). Paring this information with the functional overrepresentation results, it appears that this module contains genes that are part of the same functional pathways that moderate drinking behavior in a post-transcriptional or post-translational fashion that cannot be detected by RNA-level analyses. Given the overrepresentation of several functions involving regulation of proteins at the protein level (ligase activity, ubiquitination, mRNA modification (which could affect translation)), this module may regulate protein abundance or function in an ethanol-associated manner without affecting mRNA levels, which were measured in our expression analyses. Considering that many of the functionally overrepresented groups that did not contain hub genes were associated with neuronal signaling, synaptic processes, axon guidance, and development, these changes in protein activity likely affect short-term neuronal signaling changes and longer-term synaptic remodeling in the PFC, in response to alcohol. Sensitivity to these changes could be determined by genetic variants in the hub genes (affecting the network as a whole) or even in the ancillary genes (affecting one particular mechanism's response to change in hub gene activity).

Light Green's highly significant functional overrepresentation results revealed that its constituent genes are largely involved in ubiquitination and neurite growth. The hub genes for this module (*Cacna1b* (calcium channel, voltage-dependent, N type, alpha 1B subunit), *Cdc42bpb* (CDC42 binding protein kinase beta), *Farp1* (FERM, RhoGEF (Arhgef) and pleckstrin domain protein 1 (chondrocyte-derived)), *Tnrc6c* (trinucleotide repeat containing 6C), and *Sugp2* (SURP and G patch domain containing 2)) are not contained in the significant functional overrepresentation groups. Not much is known about the function of *Sugp2*, however the remaining four hub genes take part in well-characterized mechanisms that are highly relevant to ethanol-mediated changes in gene expression and neuroplasticity. *Cdc42bpb* regulates actin reorganization in the cytoskeleton, a process that is thought to be involved in synaptic plasticity (297, 298). *Tnrc6c* is involved in miRNA-moderated gene silencing (299). As its name suggests, *Cacna1b* encodes a subunit of voltage-gated calcium channels, which play an important role in synaptic transmission (300). Finally, *Farp1* is a guanine nucleotide exchange factor, and plays an important role in dendritic growth during motor neuron development (301). Accordingly, *Farp1* and *Cacna1b* both appeared in highly significant functional overrepresentation groups related to localization at synapses, in dendrites, and in the somatodendritic compartment. Although these genes have highly relevant functions, their expression levels were not associated with ethanol drinking traits or significantly differentially expressed genes between high drinkers and controls. However, in addition to its hub genes' relevant functionality, this module is significantly overrepresented with GSCAN GWAS-significant genes for alcohol consumption, and multiple human GWASs have revealed significant associations between several of its hub genes and substance use-relate and psychiatric traits. These associations include: *Cacna1b* with

decrease in Schizophrenia symptoms in response to paliperidone; *Cdc42bpb* with depression and smoking status; *Tnrc6c* with psychosis in Alzheimer's patients; and *Farp1* with migraines, brain structure, neuroticism, cognitive function, risk-taking behavior, response to d-amphetamines, and Alzheimer's related phenotypes (170). Much like Blue, the levels and activity of proteins encoded by the genes in this module are likely ethanol-responsive at a post-transcriptional level. As stated above, hub gene *Tnrc6c* is involved in miRNA gene silencing. This process does not reduce mRNA levels, but it prevents mRNA molecules from being translated into proteins, thereby affecting protein levels. Furthermore, this module is overrepresented with genes related to ubiquitination, a post-translational level of control over protein abundance that would also not be detected by RNAseq. The modulation of proteins encoded by genes in this module likely affects synaptic transmission (due to their localization in myelin and involvement in voltage-gated Ca^{2+} signaling) and synaptic remodeling (due to their involvement in neuronal development and cytoskeletal remodeling).

Dark Turquoise was highly significantly overrepresented with genes associated with phosphatase activity, cell-cell adhesion, and medullablastomas. This module had two hub genes: *Baiap2* (brain-specific angiogenesis inhibitor 1-associated protein 2) and *Samd4b* (sterile alpha motif domain containing 4B). Neither of these genes was present in the highly significant functional groups for this module. *Samd4b* is a transcriptional regulator of multiple apoptotic proteins, including p53, which targets *Baiap2* (302, 303). *Baiap2* mediates neurite growth via the reorganization of cytoskeletal actin, and has shown association with memory, neuroticism, and depressive symptoms in human GWASs (170, 303, 304). Like the module's Eigen Gene, these two genes' expression levels were neither different between high drinkers and controls, nor correlated with ethanol consumption or preference. However, like

Blue and Light Green, the Eigen Gene's eQTL overlapped with the Chr16 bQTL for *Last Week Mean Total Ethanol Preference*. Given its overrepresentation of phosphatase activity, Dark Turquoise is yet another example of a module potentially exhibiting ethanol-affected post-transcriptional regulation of proteins. Because genes that encode apoptotic proteins are not seen in this module, it is possible that *Samd4b* down-regulated these proteins (rendering them undetectable) in the PFC in response to ethanol, allowing *Baiap2* to mediate rerouting of neuronal pathways in response to long-term ethanol exposure, relatively unhindered.

Pale Turquoise is highly significantly overrepresented with genes in the following functional categories: location in abaxonal myelin sheath; NOTCH and Wnt signaling; abdominal wall herniation; and craniorachischisis (a severe neural tube defect). The hub gene of this module, *Atxn2l* (ataxin 2-like), is associated with intelligence and cognitive ability, as evidenced by several GWASs, but its function is not well understood (170). This gene is not significantly associated with ethanol consumption or preference, or significantly differentially expressed between high drinkers and controls in our sample. However, like Blue and Light Green, its Eigen Gene's eQTL overlapped with the Chr1 bQTL for *Last Week Mean and Whole Study Total Ethanol Consumption*. It is also not present in the highly significantly overrepresentation functional groups for this module. However, this is due to the fact that the function of this gene is unknown. Regarding the functional overrepresentation groups for this module as a whole, NOTCH and Wnt signaling involve communicating signaling cascades that mediate neurogenesis (305). Given that several genes in this module are associated with NOTCH signaling, as evidenced by the functional overrepresentation results (Table 3f), we believe that this module, and the gene *Atxn2l*, are likely involved in the process of generating novel neuronal connections, which would help

explain *Atxn2l*'s association with intelligence and cognitive ability in GWASs. This module's overlap with the Chr16 bQTL for ethanol preference and its overrepresentation for Notch and Wnt signaling pathways, and the prior associations found between its hub gene and cognition and intelligence, we believe that this module is involved in the development of novel neuronal connections (i.e. neurite outgrowth and the formation of new synapses) in response to long-term ethanol use. This relationship is likely mediated by a gene's or protein's sensitivity to ethanol due to allelic differences; or, like many of the other modules discussed here, post-trascriptional or post-translational regulation processes (that cannot be detected by examining RNA levels) resulting in protein-level changes.

Overall, this study was the first to our knowledge to characterize gene expression networks in the prefrontal cortex of Diversity Outbred mice, in the context of voluntary ethanol consumption under an intermittent access paradigm. Our gene expression analysis may have been limited by the alignment of RNAseq reads to a standard, inbred reference genome for a population of mice whose genomes contain several millions of variants that differ from this genome. This could have caused the measured expression levels of genome segments carrying many variants to be biased towards mice in which that segment was inherited from the C57BL/6J progenitor strain. This could have potentially led to the identification of false QTL for the expression of these genes, due to C57 haplotypes appearing to confer higher expression values (171). Future analyses in our laboratory will include the alignment of reads to a reference genome containing inserted SNPs unique to each mouse, based on each individual's genotype. Genotypes can be determined via the data collected from genotyping arrays described in the last chapter, or from rudimentary alignments to the C57BL/6J genome, allowing large mismatches purely for the sake of

genotyping. These programs estimate the contributing founder haplotype for each fragment by using Hidden Markov Models to determine the probability that a given read came from each of several likely regions of the genome from each possible founder strain (268) (171). However, these programs are currently not yet suitable for easy application by the public. Nonetheless, we were able to identify several networks of genes whose overall expression levels were correlated with ethanol consumption. As expected, some of these networks were overrepresented with genes that were differentially expressed between high drinking mice and controls (indicating that their expression levels were regulated by ethanol), and others were not (indicating that their expression levels moderate drinking behavior). Interestingly, the Blue module, which we believe to moderate drinking behaviors, was overrepresented with neuronal signaling involving glutamate, calcium, hormones, and neuronal growth factor, in addition to pathways involved in neuroplasticity. This was the only discussed module with such overrepresentation, and was also the only discussed module that was believed to be ethanol-regulating. The remaining modules were either ethanol-regulated (Turquoise and Brown) or were associated with ethanol via variant associations as opposed to RNA levels (Light Green, Dark Turquoise, and Pale Turquoise), and were largely overrepresented with pathways involved in protein regulation, cell cycle regulation, and neuroplasticity.

Our study was limited by the inability to collect longitudinal expression data with respect to ethanol exposure duration, rendering us unable to make definitive causal inferences regarding the relationship between ethanol consumption and gene expression. This was due to the need for collecting behavioral data related to long-term drinking on enough mice to have enough power to detect QTL for these behaviors, and could therefore not spare any animals to collect brain tissue during the early stages of ethanol exposure. Future studies

should examine the networks we identified on a longitudinal basis with respect to ethanol exposure, in order to determine whether the constituent genes are truly regulated by ethanol consumption, regulating ethanol consumption, or both. They can do so by manipulating the expression of the genes in these networks (particularly the hub genes) and measuring its effects on drinking behaviors compared to controls, and examining pre- and post-ethanol-exposure expression levels of the genes in the prefrontal cortex, in subsamples of mice from each group, at different time points with respect to experimental duration. Subsequent studies can then explore the exact molecular mechanisms through which these genes are associated with ethanol.

Additionally, this study largely focuses on the assumption that genetic associations are mediated by gene expression changes where eQTL overlap with bQTL or GWAS loci. However, it could be that the sensitivity of gene expression to the presence of ethanol is moderated by genotype. It could also be that the relationship between that specific bQTL and network is independent of their associations with drinking behaviors. Future studies should examine these relationships in a causal manner, in order to determine the nature of the associations. Such studies could examine the effects of candidate SNPs on basal gene expression and gene expression changes in response to ethanol exposure. They could also investigate the effects of candidate SNPs on drinking behaviors when their effects on candidate network expression are prevented or reversed (i.e. when the potential mediator is blocked).

Given our results, we hypothesize that gene expression networks that are not affected by alcohol but whose basal expression levels affect drinking behaviors are affiliated with the early stages of dependence development (initial consumption to tolerance development), in

which neurotransmitter levels (such as glutamate) are perturbed. The sensitivity of these systems determines the extent to which receptors are subsequently up- or down-regulated in response to ethanol-induced over- or under-stimulation. Contrastingly, networks of genes whose expression or protein levels are regulated by, or sensitive to, ethanol may be responsible for the neuroplastic changes involved in the development of withdrawal and possibly craving. In sum, this study has identified several functionally relevant groups of genes that are associated with ethanol consumption, and are likely involved in the various stages of AD development. Upon further investigation, these networks will likely reveal specific targetable genes for pharmacotherapeutic treatment of alcohol dependence.

Table 2a: Top Functional Overrepresentation Groups for Genes with Significant Differential Expression Only

Category	Name	p-value	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome
GO: Biological Process	cation transmembrane transport	4.19E-13	2.15E-09	60	738
GO: Biological Process	metal ion transport	2.63E-12	6.74E-09	64	855
GO: Biological Process	inorganic cation transmembrane transport	8.07E-12	1.38E-08	54	667
GO: Biological Process	monovalent inorganic cation transport	1.50E-11	1.92E-08	51	617
GO: Biological Process	inorganic ion transmembrane transport	3.13E-11	3.21E-08	57	754
GO: Biological Process	potassium ion transport	3.03E-10	2.59E-07	27	222
GO: Biological Process	regulation of ion transport	2.39E-09	1.75E-06	49	669
GO: Biological Process	synaptic signaling	1.57E-08	9.21E-06	48	687
GO: Biological Process	regulation of ion transmembrane transport	1.62E-08	9.21E-06	37	455
GO: Biological Process	chemical synaptic transmission	2.85E-08	1.22E-05	47	678
GO: Cellular Component	synapse	1.17E-14	6.70E-12	69	870
GO: Cellular Component	synapse part	5.03E-11	1.16E-08	54	708
GO: Cellular Component	axon	6.07E-11	1.16E-08	47	567
GO: Cellular Component	transmembrane transporter complex	3.50E-10	5.02E-08	33	327
GO: Cellular Component	transporter complex	5.60E-10	6.44E-08	33	333
GO: Cellular Component	cation channel complex	1.24E-09	1.00E-07	23	176
GO: Cellular Component	ion channel complex	1.35E-09	1.00E-07	30	291
GO: Cellular Component	axon part	1.40E-09	1.00E-07	31	309
GO: Cellular Component	somatodendritic compartment	1.01E-08	6.43E-07	56	867
GO: Cellular Component	main axon	1.91E-08	1.10E-06	14	74
GO: Molecular Function	cation channel activity	8.48E-13	1.02E-09	36	306
GO: Molecular Function	channel activity	6.92E-12	2.61E-09	44	470
GO: Molecular Function	passive transmembrane transporter activity	7.44E-12	2.61E-09	44	471
GO: Molecular Function	gated channel activity	8.67E-12	2.61E-09	36	331
GO: Molecular Function	ion channel activity	1.34E-11	3.23E-09	41	424
GO: Molecular Function	inorganic cation transmembrane transporter activity	1.91E-11	3.40E-09	47	542
GO: Molecular Function	cation transmembrane transporter activity	1.98E-11	3.40E-09	52	642
GO: Molecular Function	substrate-specific channel activity	3.73E-11	5.61E-09	41	438
GO: Molecular Function	metal ion transmembrane transporter activity	5.94E-11	7.93E-09	40	426
GO: Molecular Function	ion transmembrane transporter activity	1.50E-10	1.80E-08	61	873
Pathway	Neuronal System	1.12E-11	2.04E-08	37	351
Pathway	Calcium signaling pathway	7.95E-08	7.27E-05	21	182
Pathway	Muscle contraction	5.54E-07	3.38E-04	21	204
Pathway	Cardiac conduction	8.26E-07	3.78E-04	17	142
Pathway	Alpha adrenergic receptor signaling pathway	1.41E-06	5.15E-04	7	21
Pathway	Voltage gated Potassium channels	4.05E-06	1.24E-03	9	44
Pathway	Insulin secretion	6.27E-06	1.50E-03	12	85
Pathway	Potassium Channels	6.57E-06	1.50E-03	13	100
Pathway	Transmission across Chemical Synapses	2.18E-05	4.43E-03	19	218
Pathway	Ion homeostasis	3.18E-05	5.82E-03	9	56

Table 2b: Top Functional Overrepresentation Groups for Genes with Significant Differential Expression and Covariance with Ethanol Consumption

Category	Name	p.value	q.value FDR B.H	Hit.Count in Query List	Hit Count In Genome
GO: Biological Process	citrate metabolic process	1.53E-08	8.08E-05	10	33
GO: Biological Process	cellular respiration	3.37E-08	8.89E-05	21	178
GO: Biological Process	tricarboxylic acid cycle	6.47E-08	9.09E-05	9	29
GO: Biological Process	tricarboxylic acid metabolic process	6.89E-08	9.09E-05	10	38
GO: Biological Process	neuron projection development	1.70E-07	0.0001792	56	953
GO: Biological Process	generation of precursor metabolites and energy	2.88E-07	0.0002529	31	393
GO: Biological Process	developmental growth involved in morphogenesis	4.13E-07	0.000311	23	242
GO: Biological Process	energy derivation by oxidation of organic compounds	6.34E-07	0.0004179	25	286
GO: Biological Process	extracellular matrix organization	1.03E-06	0.000574	28	354
GO: Biological Process	extracellular structure organization	1.09E-06	0.000574	28	355
GO: Cellular Component	myelin sheath	2.55E-18	1.43E-15	35	202
GO: Cellular Component	inner mitochondrial membrane protein complex	1.15E-08	3.21E-06	17	112
GO: Cellular Component	mitochondrial protein complex	1.60E-07	2.98E-05	18	149
GO: Cellular Component	mitochondrial membrane part	2.98E-07	4.16E-05	20	189
GO: Cellular Component	mitochondrial part	5.76E-07	6.43E-05	55	987
GO: Cellular Component	mitochondrial inner membrane	5.12E-06	0.0004757	33	507
GO: Cellular Component	extracellular matrix	6.66E-06	0.0005308	30	444
GO: Cellular Component	oxidoreductase complex	1.46E-05	0.001016	12	97
GO: Cellular Component	mitochondrial envelope	1.70E-05	0.001055	41	736
GO: Cellular Component	mitochondrial membrane	2.26E-05	0.001259	39	694
GO: Molecular Function	active transmembrane transporter activity	1.27E-08	1.36E-05	32	363
GO: Molecular Function	active ion transmembrane transporter activity	2.33E-08	1.36E-05	21	175
GO: Molecular Function	anion transmembrane transporter activity	1.18E-06	0.0004599	25	297
GO: Molecular Function	ion transmembrane transporter activity	3.43E-06	0.0009481	49	873
GO: Molecular Function	glycosaminoglycan binding	4.07E-06	0.0009481	20	219
GO: Molecular Function	collagen binding	5.04E-06	0.0009785	11	72
GO: Molecular Function	amino acid transmembrane transporter activity	1.43E-05	0.00201	11	80
GO: Molecular Function	pyruvate dehydrogenase (NAD+) activity	2.03E-05	0.00201	4	7
GO: Molecular Function	pyruvate dehydrogenase activity	2.03E-05	0.00201	4	7
GO: Molecular Function	pyruvate dehydrogenase [NAD(P)+] activity	2.03E-05	0.00201	4	7
Pathway	The citric acid (TCA) cycle and respiratory electron transport	1.44E-09	2.70E-06	23	171
Pathway	Parkinson's disease	7.75E-09	7.27E-06	20	142
Pathway	Mitochondrial Electron Transport Chain	1.99E-08	1.24E-05	8	18
Pathway	Alzheimer's disease	3.87E-08	1.82E-05	21	171
Pathway	Citrate cycle (TCA cycle)	1.42E-07	5.31E-05	9	30
Pathway	Hypoacetylaspartia	4.81E-07	0.0001507	7	18
Pathway	superpathway of conversion of glucose to acetyl CoA and entry into the TCA cycle	7.94E-07	0.000213	9	36
Pathway	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	1.05E-06	0.0002453	16	126
Pathway	Huntington's disease	1.31E-06	0.0002725	20	193
Pathway	pyruvate metabolic	1.94E-06	0.000364	8	30

Table 2c: Top Functional Overrepresentation Groups for Ethanol-Correlated Only Genes

Category	Name	p-value	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome
GO: Biological Process	cellular macromolecule catabolic process	4.64E-30	4.22E-26	419	977
GO: Biological Process	proteolysis involved in cellular protein catabolic process	3.50E-25	1.59E-21	297	662
GO: Biological Process	cellular protein catabolic process	9.01E-25	2.73E-21	307	694
GO: Biological Process	proteasomal protein catabolic process	1.81E-24	4.11E-21	209	422
GO: Biological Process	modification-dependent macromolecule catabolic process	3.58E-23	6.51E-20	272	606
GO: Biological Process	modification-dependent protein catabolic process	9.95E-23	1.51E-19	268	598
GO: Biological Process	proteasome-mediated ubiquitin-dependent protein catabolic process	1.39E-22	1.81E-19	194	393
GO: Biological Process	ubiquitin-dependent protein catabolic process	2.76E-22	3.14E-19	264	590
GO: Biological Process	ribonucleoprotein complex biogenesis	1.39E-21	1.41E-18	219	468
GO: Biological Process	RNA processing	1.08E-20	9.86E-18	368	913
GO: Cellular Component	ribonucleoprotein complex	3.25E-31	3.82E-28	338	745
GO: Cellular Component	nucleolus	5.14E-24	3.02E-21	367	891
GO: Cellular Component	mitochondrial part	2.97E-21	1.16E-18	389	987
GO: Cellular Component	transferase complex	1.93E-20	5.67E-18	300	722
GO: Cellular Component	mitochondrial envelope	6.60E-17	1.55E-14	293	736
GO: Cellular Component	proteasome complex	1.64E-15	3.21E-13	54	78
GO: Cellular Component	mitochondrial membrane	8.55E-15	1.44E-12	272	694
GO: Cellular Component	ribosome	1.27E-14	1.86E-12	116	235
GO: Cellular Component	ubiquitin ligase complex	4.14E-14	5.41E-12	129	274
GO: Cellular Component	Golgi membrane	5.65E-14	6.64E-12	276	716
GO: Molecular Function	ubiquitin protein ligase activity	4.77E-14	1.21E-10	107	212
GO: Molecular Function	ubiquitin-like protein ligase activity	1.20E-13	1.52E-10	108	217
GO: Molecular Function	ubiquitin-protein transferase activity	1.84E-11	1.55E-08	171	414
GO: Molecular Function	GDP binding	3.21E-11	1.79E-08	40	59
GO: Molecular Function	ubiquitin-like protein transferase activity	3.54E-11	1.79E-08	179	441
GO: Molecular Function	GTPase activity	9.14E-11	3.85E-08	108	236
GO: Molecular Function	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	1.28E-10	4.37E-08	300	829
GO: Molecular Function	hydrolase activity, acting on acid anhydrides	1.49E-10	4.37E-08	300	830
GO: Molecular Function	nucleoside-triphosphatase activity	1.55E-10	4.37E-08	286	785
GO: Molecular Function	pyrophosphatase activity	2.51E-10	6.36E-08	298	827
Pathway	Infectious disease	9.43E-20	2.97E-16	184	393
Pathway	Membrane Trafficking	1.66E-18	2.61E-15	256	614
Pathway	Vesicle-mediated transport	1.36E-16	1.43E-13	265	660
Pathway	Antigen processing: Ubiquitination & Proteasome degradation	2.00E-16	1.58E-13	148	314
Pathway	Cell Cycle	2.88E-15	1.81E-12	249	624
Pathway	Oxygen-dependent proline hydroxylation of Hypoxia-inducible Factor Alpha	4.85E-15	2.40E-12	49	69
Pathway	HIV Infection	5.34E-15	2.40E-12	118	240
Pathway	Processing of Capped Intron-Containing Pre-mRNA	1.38E-14	5.44E-12	120	248
Pathway	mRNA Splicing - Major Pathway	3.23E-14	1.04E-11	97	188
Pathway	mRNA Splicing	3.29E-14	1.04E-11	100	196

Top highly significant ($q < 0.1$) functional overrepresentation groups from TopFunn ontological analysis, for genes that were significant in DEseq2 high drinker v. control analysis only (a), significant in both DEseq2 analysis and in regression testing for covariance with ethanol consumption (b), and significant only in regression testing (c). Category and Name columns represent the functional categories selected in TopFunn and the name of the exact function or related biological entity for which the module was overrepresented, respectively. Hit Count in Query List indicates the number of genes in the module that were identified in the group of total genes, the size of which is indicated by Hit Count in Genome.

Table 3a: Top Turquoise Functional Overrepresentation Groups

Category	Name	p-value	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome
GO: Molecular Function	ubiquitin-like protein transferase activity	1.11E-12	2.36E-09	131	441
GO: Molecular Function	ubiquitin-protein transferase activity	2.41E-12	2.55E-09	124	414
GO: Molecular Function	GDP binding	1.74E-10	1.23E-07	31	59
GO: Molecular Function	ubiquitin-like protein ligase activity	5.03E-09	2.66E-06	70	217
GO: Molecular Function	ubiquitin protein ligase activity	2.64E-08	1.12E-05	67	212
GO: Molecular Function	ubiquitin-like protein ligase binding	9.59E-07	3.20E-04	77	277
GO: Molecular Function	ubiquitin protein ligase binding	1.06E-06	3.20E-04	76	273
GO: Molecular Function	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	4.76E-06	1.21E-03	184	829
GO: Molecular Function	hydrolase activity, acting on acid anhydrides	5.14E-06	1.21E-03	184	830
GO: Molecular Function	pyrophosphatase activity	6.14E-06	1.30E-03	183	827
GO: Molecular Function	ubiquitin-like protein binding	6.99E-06	1.35E-03	41	127
GO: Molecular Function	ubiquitin-like protein conjugating enzyme activity	1.08E-05	1.91E-03	16	32
GO: Molecular Function	ligase activity	1.58E-05	2.57E-03	101	415
GO: Molecular Function	ubiquitin conjugating enzyme activity	3.37E-05	5.09E-03	15	31
GO: Molecular Function	nucleoside-triphosphatase activity	4.56E-05	6.43E-03	170	785
GO: Molecular Function	ligase activity, forming carbon-nitrogen bonds	5.83E-05	7.11E-03	74	293
GO: Molecular Function	ribonucleoprotein complex binding	5.88E-05	7.11E-03	37	120
GO: Molecular Function	RNA-dependent ATPase activity	6.62E-05	7.11E-03	24	66
GO: Molecular Function	ATP-dependent RNA helicase activity	6.62E-05	7.11E-03	24	66
GO: Molecular Function	GTPase activity	6.72E-05	7.11E-03	62	236
GO: Biological Process	protein modification by small protein conjugation	8.75E-20	7.12E-16	254	903
GO: Biological Process	protein ubiquitination	4.57E-16	1.86E-12	216	781
GO: Biological Process	RNA processing	3.95E-14	1.07E-10	237	913
GO: Biological Process	modification-dependent macromolecule catabolic process	3.78E-13	7.50E-10	169	606
GO: Biological Process	modification-dependent protein catabolic process	4.61E-13	7.50E-10	167	598
GO: Biological Process	ubiquitin-dependent protein catabolic process	5.61E-13	7.61E-10	165	590
GO: Biological Process	proteolysis involved in cellular protein catabolic process	7.46E-13	8.67E-10	180	662
GO: Biological Process	protein polyubiquitination	8.68E-13	8.82E-10	88	254
GO: Biological Process	cellular protein-containing complex assembly	1.33E-12	1.20E-09	247	991
GO: Biological Process	cellular protein catabolic process	2.41E-12	1.96E-09	185	694
GO: Biological Process	mRNA processing	4.53E-12	3.35E-09	138	479
GO: Biological Process	proteasomal protein catabolic process	1.33E-10	8.98E-08	121	422
GO: Biological Process	protein catabolic process	1.46E-10	9.14E-08	209	844
GO: Biological Process	RNA splicing	2.09E-10	1.15E-07	116	402
GO: Biological Process	proteasome-mediated ubiquitin-dependent protein catabolic process	2.11E-10	1.15E-07	114	393
GO: Biological Process	RNA splicing, via transesterification reactions	2.36E-10	1.20E-07	95	309
GO: Biological Process	mitotic cell cycle process	5.23E-10	2.50E-07	224	931
GO: Biological Process	mRNA splicing, via spliceosome	1.34E-09	5.75E-07	92	305
GO: Biological Process	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	1.34E-09	5.75E-07	92	305
GO: Biological Process	ribonucleoprotein complex biogenesis	8.13E-09	3.31E-06	125	468
GO: Cellular Component	transferase complex	2.51E-20	2.61E-17	213	722
GO: Cellular Component	nucleolus	1.57E-15	8.17E-13	234	891
GO: Cellular Component	nucleoplasm part	3.17E-11	1.10E-08	186	732
GO: Cellular Component	ubiquitin ligase complex	1.93E-10	5.02E-08	86	274
GO: Cellular Component	Golgi membrane	1.73E-09	3.60E-07	176	716
GO: Cellular Component	mitochondrial matrix	4.71E-09	8.18E-07	115	425
GO: Cellular Component	mitochondrial part	7.40E-09	1.02E-06	226	987
GO: Cellular Component	ribonucleoprotein complex	7.81E-09	1.02E-06	179	745
GO: Cellular Component	Golgi apparatus part	1.67E-08	1.93E-06	213	928
GO: Cellular Component	nuclear body	2.06E-08	2.15E-06	100	364
GO: Cellular Component	vacuolar membrane	1.18E-07	1.11E-05	144	593
GO: Cellular Component	endoplasmic reticulum membrane	3.13E-07	2.72E-05	219	994
GO: Cellular Component	cullin-RING ubiquitin ligase complex	3.67E-07	2.94E-05	50	154
GO: Cellular Component	microbody	4.33E-07	3.01E-05	47	142
GO: Cellular Component	peroxisome	4.33E-07	3.01E-05	47	142
GO: Cellular Component	transport vesicle	6.34E-07	4.13E-05	97	374
GO: Cellular Component	serine/threonine protein kinase complex	1.23E-06	7.54E-05	31	81
GO: Cellular Component	transferase complex, transferring phosphorus-containing groups	2.08E-06	1.21E-04	68	244
GO: Cellular Component	endosomal part	2.33E-06	1.28E-04	107	434
GO: Cellular Component	vacuolar part	3.43E-06	1.79E-04	158	700

Human Phenotype	Generalized hypotonia	1.09E-09	5.59E-06	188	808
Human Phenotype	Microcephaly	5.35E-09	9.15E-06	191	839
Human Phenotype	Decreased head circumference	5.35E-09	9.15E-06	191	839
Human Phenotype	Delayed speech and language development	1.07E-07	1.37E-04	104	407
Human Phenotype	Nystagmus	2.27E-07	2.00E-04	179	811
Human Phenotype	Abnormality of the cerebral white matter	2.36E-07	2.00E-04	133	563
Human Phenotype	Abnormal involuntary eye movements	2.73E-07	2.00E-04	181	824
Human Phenotype	Neurological speech impairment	8.20E-07	5.16E-04	177	814
Human Phenotype	Aplasia/Hypoplasia of the corpus callosum	1.15E-06	5.16E-04	112	467
Human Phenotype	Strabismus	1.20E-06	5.16E-04	157	708
Human Phenotype	Spasticity	1.21E-06	5.16E-04	185	863
Human Phenotype	Hypertonia	1.21E-06	5.16E-04	185	863
Human Phenotype	Abnormal corpus callosum morphology	1.50E-06	5.93E-04	115	485
Human Phenotype	Abnormality of the cerebral subcortex	2.39E-06	8.75E-04	143	639
Human Phenotype	Abnormal conjugate eye movement	1.09E-05	3.72E-03	157	733
Human Phenotype	Feeding difficulties	1.33E-05	4.26E-03	137	625
Human Phenotype	Intellectual disability, severe	2.69E-05	7.89E-03	66	258
Human Phenotype	Abnormal location of ears	2.77E-05	7.89E-03	121	546
Human Phenotype	Open mouth	3.51E-05	9.48E-03	31	95
Human Phenotype	Motor delay	4.00E-05	1.03E-02	89	379
Mouse Phenotype	embryonic lethality prior to organogenesis	1.81E-13	1.24E-09	219	792
Mouse Phenotype	embryonic lethality prior to tooth bud stage	1.64E-12	5.62E-09	236	887
Mouse Phenotype	embryonic lethality between implantation and placentation	1.31E-07	2.99E-04	122	447
Mouse Phenotype	abnormal myocardial fiber physiology	1.43E-05	2.45E-02	53	172
Mouse Phenotype	embryonic growth retardation	2.69E-05	3.68E-02	152	639
Mouse Phenotype	prenatal growth retardation	5.07E-05	5.79E-02	170	736
Mouse Phenotype	abnormal learning/memory/conditioning	7.41E-05	6.86E-02	156	671
Mouse Phenotype	abnormal cognition	8.02E-05	6.86E-02	156	672
Mouse Phenotype	abnormal cell cycle	1.31E-04	9.99E-02	85	332
Pathway	Membrane Trafficking	2.27E-10	6.61E-07	157	614
Pathway	Antigen processing: Ubiquitination & Proteasome degradation	5.31E-10	7.75E-07	93	314
Pathway	Ubiquitin mediated proteolysis	3.34E-09	3.25E-06	50	137
Pathway	Class I MHC mediated antigen processing & presentation	7.20E-09	5.25E-06	103	376
Pathway	Cell Cycle	9.01E-09	5.26E-06	153	624
Pathway	Vesicle-mediated transport	1.66E-08	8.07E-06	159	660
Pathway	Processing of Capped Intron-Containing Pre-mRNA	2.21E-08	9.21E-06	74	248
Pathway	mRNA Splicing	1.75E-07	6.40E-05	60	196
Pathway	S Phase	2.07E-07	6.70E-05	45	132
Pathway	Cell Cycle, Mitotic	2.77E-07	8.08E-05	126	517
Pathway	mRNA Splicing - Major Pathway	5.04E-07	1.34E-04	57	188
Pathway	Processing and activation of SUMO	5.64E-07	1.36E-04	9	10
Pathway	Mitotic G1-G1/S phases	6.05E-07	1.36E-04	47	145
Pathway	Mitotic Anaphase	7.35E-07	1.53E-04	55	181
Pathway	Mitotic Metaphase and Anaphase	8.93E-07	1.74E-04	55	182
Pathway	Separation of Sister Chromatids	1.17E-06	2.14E-04	52	170
Pathway	Retrograde transport at the Trans-Golgi-Network	1.58E-06	2.71E-04	22	49
Pathway	GABA A receptor activation	1.87E-06	3.03E-04	10	13
Pathway	Organelle biogenesis and maintenance	2.86E-06	4.39E-04	87	341
Pathway	RNA transport	3.37E-06	4.92E-04	51	171
Disease	Muscle hypotonia	9.55E-09	3.86E-05	134	575
Disease	Global developmental delay	9.79E-09	3.86E-05	167	757
Disease	Epilepsy	2.85E-08	7.48E-05	201	962
Disease	Mental Retardation	6.25E-08	1.23E-04	186	885
Disease	Poor school performance	6.22E-07	6.99E-04	140	649
Disease	Low intelligence	6.22E-07	6.99E-04	140	649
Disease	Dull intelligence	6.22E-07	6.99E-04	140	649
Disease	Mental deficiency	7.46E-07	7.34E-04	140	651
Disease	Hyperreflexia	1.18E-06	1.04E-03	57	209
Disease	Cognitive delay	2.62E-06	1.77E-03	131	614
Disease	Mental and motor retardation	2.62E-06	1.77E-03	131	614
Disease	Mental Retardation, X-Linked	2.69E-06	1.77E-03	35	109
Disease	Leigh Disease	9.63E-06	5.83E-03	30	92
Disease	Dysarthria	1.76E-05	9.92E-03	51	196
Disease	Acidosis, Lactic	1.95E-05	1.03E-02	40	142
Disease	Action Tremor	2.71E-05	1.28E-02	14	30
Disease	Muscle Spasticity	2.76E-05	1.28E-02	61	251
Disease	Drooling	3.47E-05	1.52E-02	13	27
Disease	Widely spaced teeth	4.27E-05	1.77E-02	14	31
Disease	Small head	4.64E-05	1.83E-02	83	375

Table 3a: Top Brown Functional Overrepresentation Groups

Category	Name	p-value	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome
GO: Biological Process	response to UV	8.09E-06	3.15E-02	14	140
GO: Biological Process	activation of cysteine-type endopeptidase activity involved in apoptotic process by cytochrome c	2.22E-05	4.32E-02	4	8
GO: Biological Process	UV-damage excision repair	6.40E-05	6.58E-02	4	10
GO: Biological Process	DNA repair	6.76E-05	6.58E-02	28	516
GO: Biological Process	DNA recombination	9.86E-05	7.68E-02	18	267
GO: Cellular Component	transferase complex	5.93E-05	2.74E-02	35	722
GO: Cellular Component	INO80-type complex	1.11E-04	2.74E-02	5	21
GO: Cellular Component	Ino80 complex	2.63E-04	4.31E-02	4	14
GO: Cellular Component	DNA helicase complex	3.52E-04	4.33E-02	4	15
GO: Cellular Component	ubiquitin ligase complex	9.33E-04	9.18E-02	16	274
Human Phenotype	Increased sensitivity to ionizing radiation	1.29E-06	3.31E-03	8	32
Human Phenotype	Abnormality of DNA repair	2.14E-06	3.31E-03	8	34
Human Phenotype	Abnormality of the cell cycle	7.90E-06	8.16E-03	8	40
Human Phenotype	Multiple cutaneous malignancies	3.15E-05	1.95E-02	4	8
Human Phenotype	Flat nasal alae	3.15E-05	1.95E-02	4	8
Human Phenotype	Abnormality of chromosome stability	5.11E-05	2.36E-02	8	51
Human Phenotype	Severe photosensitivity	6.07E-05	2.36E-02	5	17
Human Phenotype	Poikiloderma	6.09E-05	2.36E-02	6	27
Human Phenotype	Microcephaly	8.76E-05	2.71E-02	40	839
Human Phenotype	Decreased head circumference	8.76E-05	2.71E-02	40	839
Human Phenotype	Skeletal muscle atrophy	1.36E-04	3.82E-02	27	483
Human Phenotype	Intellectual disability, progressive	1.48E-04	3.82E-02	13	149
Human Phenotype	Conjunctival telangiectasia	2.90E-04	6.42E-02	5	23
Human Phenotype	Ankyloblepharon	2.90E-04	6.42E-02	4	13
Human Phenotype	Abnormal vasculature of the conjunctiva morphology	4.39E-04	8.67E-02	5	25
Human Phenotype	Complement deficiency	4.48E-04	8.67E-02	8	69
Pathway	Global Genome Nucleotide Excision Repair (GG-NER)	4.35E-07	4.91E-04	12	85
Pathway	Cytochrome c-mediated apoptotic response	1.27E-06	4.91E-04	4	5
Pathway	Activation of caspases through apoptosome-mediated cleavage	1.27E-06	4.91E-04	4	5
Pathway	Nucleotide Excision Repair	1.44E-06	4.91E-04	13	112
Pathway	DNA Repair	3.52E-06	9.62E-04	22	319
Pathway	Apoptotic factor-mediated response	8.55E-06	1.95E-03	4	7
Pathway	Nucleotide excision repair	8.22E-05	1.52E-02	7	47
Pathway	SMAC binds to IAPs	1.11E-04	1.52E-02	3	5
Pathway	SMAC-mediated apoptotic response	1.11E-04	1.52E-02	3	5
Pathway	SMAC-mediated dissociation of IAP:caspase complexes	1.11E-04	1.52E-02	3	5
Pathway	DNA Damage Recognition in GG-NER	2.23E-04	2.77E-02	6	39
Pathway	Resolution of D-loop Structures through Synthesis-Dependent Strand Annealing (SDSA)	3.09E-04	3.52E-02	5	27
Pathway	Formation of Incision Complex in GG-NER	3.86E-04	4.06E-02	6	43
Pathway	intrinsic apoptotic	4.37E-04	4.27E-02	5	29
Pathway	Apoptosis is mediated by caspases, cysteine proteases arranged in a proteolytic cascade.	7.63E-04	6.96E-02	4	19
Pathway	Resolution of D-loop Structures through Holliday Junction Intermediates	1.07E-03	9.14E-02	5	35
Pathway	Role of Mitochondria in Apoptotic Signaling	1.14E-03	9.15E-02	4	21
Pathway	Resolution of D-Loop Structures	1.22E-03	9.27E-02	5	36
Pathway	Apoptotic Signaling in Response to DNA Damage	1.37E-03	9.83E-02	4	22
Disease	Progressive mental retardation	8.30E-06	2.65E-02	7	37
Disease	Cockayne Syndrome	9.47E-05	9.19E-02	8	71
Disease	Cystic Kidney Diseases	1.05E-04	9.19E-02	5	24
Disease	Conjunctival telangiectasis	1.44E-04	9.19E-02	4	14
Disease	Nephronophthisis, familial juvenile	1.44E-04	9.19E-02	4	14

Table 3c: Top Blue Functional Overrepresentation Groups

Category	Name	p-value	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome
GO: Molecular Function	GTPase binding	1.36E-13	2.26E-10	67	327
GO: Molecular Function	cytoskeletal protein binding	6.69E-13	4.85E-10	130	886
GO: Molecular Function	small GTPase binding	8.76E-13	4.85E-10	62	301
GO: Molecular Function	Ras GTPase binding	4.38E-12	1.82E-09	58	281
GO: Molecular Function	GTPase regulator activity	3.45E-10	1.15E-07	58	312
GO: Molecular Function	GTPase activator activity	1.51E-09	4.18E-07	53	283
GO: Molecular Function	chromatin binding	2.95E-09	7.00E-07	79	513
GO: Molecular Function	nucleoside-triphosphatase regulator activity	5.88E-09	1.22E-06	59	344
GO: Molecular Function	steroid hormone receptor binding	9.38E-09	1.73E-06	25	89
GO: Molecular Function	protein domain specific binding	1.33E-08	2.20E-06	100	729
GO: Molecular Function	Rab GTPase binding	1.45E-07	2.18E-05	29	129
GO: Molecular Function	enzyme activator activity	1.71E-07	2.37E-05	74	515
GO: Molecular Function	estrogen receptor binding	2.37E-07	3.02E-05	16	47
GO: Molecular Function	calmodulin binding	4.65E-07	5.51E-05	36	190
GO: Molecular Function	hormone receptor binding	5.30E-07	5.87E-05	35	183
GO: Molecular Function	ion channel binding	1.24E-06	1.22E-04	27	127
GO: Molecular Function	protein kinase binding	1.25E-06	1.22E-04	82	620
GO: Molecular Function	kinase binding	1.37E-06	1.26E-04	89	691
GO: Molecular Function	nuclear hormone receptor binding	2.00E-06	1.75E-04	31	161
GO: Molecular Function	beta-catenin binding	3.26E-06	2.70E-04	21	89
GO: Biological Process	cell projection morphogenesis	3.66E-14	1.19E-10	136	904
GO: Biological Process	neuron projection development	4.57E-14	1.19E-10	141	953
GO: Biological Process	endomembrane system organization	5.21E-14	1.19E-10	99	578
GO: Biological Process	cell part morphogenesis	9.99E-14	1.71E-10	137	925
GO: Biological Process	histone modification	1.94E-13	2.66E-10	82	447
GO: Biological Process	covalent chromatin modification	4.48E-13	5.11E-10	83	462
GO: Biological Process	neuron projection morphogenesis	1.93E-12	1.89E-09	101	630
GO: Biological Process	regulation of GTPase activity	6.65E-12	5.32E-09	106	688
GO: Biological Process	cell morphogenesis involved in neuron differentiation	6.99E-12	5.32E-09	95	590
GO: Biological Process	Golgi vesicle transport	9.91E-12	6.79E-09	65	340
GO: Biological Process	positive regulation of GTPase activity	2.97E-11	1.85E-08	98	632
GO: Biological Process	regulation of cell projection organization	5.49E-11	3.14E-08	100	657
GO: Biological Process	vacuolar transport	1.38E-10	7.28E-08	56	288
GO: Biological Process	regulation of cellular component biogenesis	1.61E-10	7.89E-08	117	829
GO: Biological Process	regulation of cell morphogenesis	4.95E-10	2.26E-07	92	609
GO: Biological Process	positive regulation of cell projection organization	7.85E-10	3.24E-07	64	367
GO: Biological Process	brain development	8.03E-10	3.24E-07	107	755
GO: Biological Process	chromatin organization	1.21E-09	4.50E-07	108	770
GO: Biological Process	axonogenesis	1.25E-09	4.50E-07	76	475
GO: Biological Process	head development	1.53E-09	5.25E-07	111	802
GO: Cellular Component	postsynaptic specialization	3.33E-20	1.41E-17	64	232
GO: Cellular Component	postsynaptic density	3.33E-20	1.41E-17	64	232
GO: Cellular Component	synapse	7.50E-20	2.12E-17	145	870
GO: Cellular Component	postsynapse	3.33E-18	5.79E-16	91	449
GO: Cellular Component	excitatory synapse	3.42E-18	5.79E-16	64	252
GO: Cellular Component	synapse part	1.48E-17	2.09E-15	121	708
GO: Cellular Component	dendrite	4.93E-14	5.97E-12	99	590
GO: Cellular Component	somatodendritic compartment	1.79E-13	1.90E-11	128	867
GO: Cellular Component	dendritic spine	7.23E-12	6.82E-10	39	151
GO: Cellular Component	neuron spine	1.12E-11	9.53E-10	39	153
GO: Cellular Component	axon	5.67E-09	4.37E-07	83	567
GO: Cellular Component	transferase complex	1.29E-08	8.50E-07	98	722
GO: Cellular Component	nucleoplasm part	1.30E-08	8.50E-07	99	732
GO: Cellular Component	perinuclear region of cytoplasm	1.48E-08	8.98E-07	98	724
GO: Cellular Component	cell leading edge	1.20E-07	6.76E-06	60	389
GO: Cellular Component	site of polarized growth	1.99E-07	1.05E-05	36	187
GO: Cellular Component	microtubule organizing center	2.41E-07	1.20E-05	86	646
GO: Cellular Component	axon part	3.08E-07	1.45E-05	50	309
GO: Cellular Component	endosome	4.02E-07	1.79E-05	103	826
GO: Cellular Component	neuronal cell body	4.61E-07	1.96E-05	82	616

Human Phenotype	Brain atrophy	3.23E-09	8.10E-06	83	488
Human Phenotype	Brain very small	3.57E-09	8.10E-06	83	489
Human Phenotype	Cerebral hypoplasia	1.11E-08	1.49E-05	87	534
Human Phenotype	Microcephaly	1.64E-08	1.49E-05	121	839
Human Phenotype	Decreased head circumference	1.64E-08	1.49E-05	121	839
Human Phenotype	Cerebral atrophy	2.11E-08	1.59E-05	77	457
Human Phenotype	Atrophy/Degeneration affecting the cerebrum	4.91E-08	3.18E-05	78	474
Human Phenotype	Abnormal corpus callosum morphology	6.28E-08	3.56E-05	79	485
Human Phenotype	Abnormality of the cerebral subcortex	1.15E-07	5.78E-05	96	639
Human Phenotype	Aplasia/Hypoplasia of the corpus callosum	2.66E-07	1.10E-04	75	467
Human Phenotype	Abnormality of globe location	2.75E-07	1.10E-04	109	768
Human Phenotype	Hyperactivity	2.92E-07	1.10E-04	49	258
Human Phenotype	Delayed speech and language development	1.11E-06	3.87E-04	66	407
Human Phenotype	Abnormality of the cerebral white matter	1.20E-06	3.90E-04	84	563
Human Phenotype	Abnormality of globe location or size	1.36E-06	3.94E-04	124	933
Human Phenotype	Abnormality of the chin	1.46E-06	3.94E-04	44	233
Human Phenotype	Brachycephaly	1.48E-06	3.94E-04	40	203
Human Phenotype	Spasticity	1.95E-06	4.65E-04	116	863
Human Phenotype	Hypertonia	1.95E-06	4.65E-04	116	863
Human Phenotype	Hyperreflexia	2.24E-06	5.08E-04	87	598
Mouse Phenotype	abnormal CNS synaptic transmission	5.02E-10	2.99E-06	121	734
Mouse Phenotype	abnormal synaptic transmission	1.99E-08	5.91E-05	143	965
Mouse Phenotype	abnormal prenatal body size	3.18E-07	4.79E-04	137	956
Mouse Phenotype	reduced long term potentiation	3.68E-07	4.79E-04	36	156
Mouse Phenotype	abnormal emotion/affect behavior	4.58E-07	4.79E-04	99	638
Mouse Phenotype	abnormal cerebral hemisphere morphology	4.82E-07	4.79E-04	100	647
Mouse Phenotype	abnormal telencephalon morphology	5.70E-07	4.86E-04	126	870
Mouse Phenotype	abnormal long term potentiation	9.00E-07	6.71E-04	47	238
Mouse Phenotype	preweaning lethality, incomplete penetrance	1.10E-06	7.30E-04	99	650
Mouse Phenotype	abnormal spatial learning	1.43E-06	8.51E-04	48	249
Mouse Phenotype	abnormal motor coordination/balance	1.67E-06	9.03E-04	124	870
Mouse Phenotype	abnormal reflex	1.86E-06	9.23E-04	127	898
Mouse Phenotype	abnormal cognition	2.83E-06	1.21E-03	100	672
Mouse Phenotype	abnormal neurite morphology	2.91E-06	1.21E-03	75	465
Mouse Phenotype	abnormal hippocampus morphology	3.06E-06	1.21E-03	63	370
Mouse Phenotype	abnormal brain development	3.27E-06	1.21E-03	118	829
Mouse Phenotype	abnormal prepulse inhibition	3.52E-06	1.21E-03	46	242
Mouse Phenotype	small cerebellum	3.67E-06	1.21E-03	35	164
Mouse Phenotype	abnormal learning/memory/conditioning	4.66E-06	1.43E-03	99	671
Mouse Phenotype	abnormal cerebellum morphology	5.10E-06	1.43E-03	75	472
Pathway	Membrane Trafficking	9.55E-11	2.22E-07	92	614
Pathway	Vesicle-mediated transport	4.35E-09	5.04E-06	92	660
Pathway	Phosphatidylinositol signaling system	7.45E-09	5.76E-06	26	97
Pathway	Neuronal System	2.47E-08	1.43E-05	57	351
Pathway	Reelin signaling pathway	5.18E-08	2.40E-05	13	29
Pathway	Chromatin modifying enzymes	1.39E-07	4.60E-05	47	279
Pathway	Chromatin organization	1.39E-07	4.60E-05	47	279
Pathway	Transmission across Chemical Synapses	3.33E-07	9.64E-05	39	218
Pathway	Axon guidance	8.44E-07	2.17E-04	74	554
Pathway	Thyroid hormone signaling pathway	1.36E-06	3.16E-04	25	116
Pathway	Transport to the Golgi and subsequent modification	1.82E-06	3.84E-04	31	165
Pathway	RORA activates gene expression	2.78E-06	5.38E-04	11	28
Pathway	Signalling by NGF	6.20E-06	1.11E-03	64	483
Pathway	Control of Gene Expression by Vitamin D Receptor	7.73E-06	1.28E-03	7	12
Pathway	Asparagine N-linked glycosylation	9.19E-06	1.32E-03	43	285
Pathway	Alpha adrenergic receptor signaling pathway	9.83E-06	1.32E-03	9	21
Pathway	Rho GTPase cycle	9.88E-06	1.32E-03	27	145
Pathway	CXCR4-mediated signaling events	1.02E-05	1.32E-03	18	76
Pathway	Endocytosis	1.17E-05	1.36E-03	40	260

Table 3d: Top Light Green Functional Overrepresentation Groups

Category	Name	p-value	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome
GO: Molecular Function	ubiquitin-protein transferase activity	2.22E-04	8.80E-02	9	414
GO: Molecular Function	ubiquitin-like protein transferase activity	3.53E-04	8.80E-02	9	441
GO: Biological Process	regulation of cellular response to heat	3.89E-05	6.70E-02	5	76
GO: Cellular Component	somatodendritic compartment	2.91E-04	4.19E-02	13	867
GO: Cellular Component	dendrite	6.02E-04	4.19E-02	10	590
GO: Cellular Component	growth cone membrane	6.47E-04	4.19E-02	2	8
GO: Cellular Component	intraciliary transport particle A	6.47E-04	4.19E-02	2	8
GO: Cellular Component	neuronal cell body	8.38E-04	4.34E-02	10	616
GO: Cellular Component	ciliary tip	1.38E-03	5.78E-02	3	45
GO: Cellular Component	cell body	1.98E-03	5.78E-02	10	691
GO: Cellular Component	growth cone	1.98E-03	5.78E-02	5	182
GO: Cellular Component	axon terminus	2.18E-03	5.78E-02	5	186
GO: Cellular Component	site of polarized growth	2.23E-03	5.78E-02	5	187
GO: Cellular Component	neuron projection terminus	2.92E-03	6.87E-02	5	199
GO: Cellular Component	synapse	3.37E-03	7.00E-02	11	870
GO: Cellular Component	histone deacetylase complex	3.61E-03	7.00E-02	3	63
GO: Cellular Component	pericentric heterochromatin	3.82E-03	7.00E-02	2	19
GO: Cellular Component	axon part	4.05E-03	7.00E-02	6	309
GO: Cellular Component	COPI-coated vesicle	5.57E-03	9.02E-02	2	23
GO: Cellular Component	axon	6.50E-03	9.90E-02	8	567
Mouse Phenotype	abnormal lung position or orientation	3.68E-05	6.72E-02	5	58
Pathway	Intra-Golgi and retrograde Golgi-to-ER traffic	1.18E-05	1.36E-03	31	180
Disease	Mental Retardation	2.54E-18	1.58E-14	137	885
Disease	Autistic Disorder	3.47E-16	1.09E-12	106	644
Disease	Poor school performance	1.87E-13	2.33E-10	100	649
Disease	Low intelligence	1.87E-13	2.33E-10	100	649
Disease	Dull intelligence	1.87E-13	2.33E-10	100	649
Disease	Mental deficiency	2.27E-13	2.36E-10	100	651
Disease	Dysarthria	7.64E-12	6.81E-09	44	196
Disease	Seizures	1.66E-11	1.29E-08	128	982
Disease	Muscle hypotonia	2.05E-11	1.42E-08	87	575
Disease	Global developmental delay	1.67E-10	1.04E-07	103	757
Disease	Small head	5.01E-10	2.84E-07	62	375
Disease	Delayed speech and language development	2.04E-09	9.11E-07	29	116
Disease	Speech impairment	2.04E-09	9.11E-07	29	116
Disease	Language Delay	2.04E-09	9.11E-07	29	116
Disease	Cognitive delay	3.14E-09	1.23E-06	85	614
Disease	Mental and motor retardation	3.14E-09	1.23E-06	85	614
Disease	Cerebellar Ataxia	4.79E-09	1.76E-06	68	453
Disease	Speech Delay	5.60E-09	1.94E-06	30	128
Disease	Degenerative brain disorder	7.31E-09	2.40E-06	27	108
Disease	Epilepsy	8.67E-09	2.71E-06	117	962

Table 3e: Top Dark Turquoise Functional Overrepresentation Groups

Category	Name	p-value	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome
GO: Molecular Function	phosphatase inhibitor activity	4.47E-04	8.15E-02	3	41
GO: Molecular Function	phosphatase binding	5.99E-04	8.15E-02	5	187
GO: Molecular Function	cell-cell adhesion mediator activity	8.43E-04	8.15E-02	2	12
Disease	Medulloblastoma	2.25E-05	2.27E-02	9	558

Table 3f: Top Pale Turquoise Functional Overrepresentation Groups

Category	Name	p-value	q-value FDR B&H	Hit Count in Query List	Hit Count in Genome
GO: Cellular Component	myelin sheath abaxonal region	2.30E-04	3.32E-02	2	9
Mouse Phenotype	herniated abdominal wall	2.18E-05	2.00E-02	3	20
Pathway	Notch signaling pathway	2.03E-04	3.87E-02	3	48
Pathway	Constitutive Signaling by NOTCH1 HD Domain Mutants	5.78E-04	3.87E-02	2	15
Pathway	the planar cell polarity Wnt signaling	5.78E-04	3.87E-02	2	15
Pathway	Signaling by NOTCH1 HD Domain Mutants in Cancer	5.78E-04	3.87E-02	2	15
Pathway	NOTCH2 Activation and Transmission of Signal to the Nucleus	1.15E-03	6.14E-02	2	21
Disease	Craniorachischisis	1.81E-04	9.29E-02	2	9

Top highly significant ($q < 0.1$) functional overrepresentation groups from TopFunn ontological analysis, for modules Turquoise (a), Brown (b), Blue (c), Light Green (d), Dark Turquoise (e), and Pale Turquoise (f). Category and Name columns represent the functional categories selected in TopFunn and the name of the exact function or related biological entity for which the module was overrepresented, respectively. Hit Count in Query List indicates the number of genes in the module that were identified in the group of total genes, the size of which is indicated by Hit Count in Genome.

Chapter 6

Final Discussion: Overview and Implications

Overview of Findings and Implications

Despite the estimated 50% heritability of alcohol use disorder (21, 22), genetic association studies have yet to account for a large proportion of this variability (23). Well-powered studies with extremely large sample sizes (hundreds of thousands to over a million) have identified hundreds of loci associated with alcohol consumption, but one recent study has indicated that problematic drinking may be more closely genetically correlated with other psychiatric disorders than it is with alcohol consumption (25). Furthermore, although these studies are able to identify single genes with individually small effect sizes, they do not provide information regarding larger biological frameworks related to the trait. Gene expression studies are capable of providing this insight into functional networks through which these loci are likely affecting drinking behaviors. Mouse models are important resources for examining gene expression patterns, because they provide experimental control not available in human studies. The benefit of human participants is that they present fully complex behaviors generalizable to non-experimental settings, and possess genomes with millions of testable variants with small linkage disequilibrium blocks, allowing for more precise mapping of affective loci than that which can be done in mouse models.

The studies presented in this thesis capitalize on the best of both worlds by examining gene expression networks in mouse brains, in the context of acute ethanol treatment and voluntary ethanol consumption, and human genetic association data. We successfully identified functionally interrelated groups of genes associated with ethanol exposure and ethanol consumption in genetically complex mice. Many of these networks were significantly overrepresented with genes associated with alcohol consumption or dependence in human GWAS samples. Furthermore, we identified genetic loci associated with ethanol consumption

behaviors in the Diversity Outbred stock of mice, with unprecedented genomic precision, and replicated a finding from a large GWAS on average drinking levels. Finally, some of the identified loci overlapped with loci containing variants associated with expression levels of ethanol-associated networks. These networks and loci were largely associated with ubiquitin activity, neurite outgrowth, and axonal guidance, indicating that neuroplastic mechanisms are involved in drinking behaviors well before the development of withdrawal or craving symptoms arise. Experiments and implications are discussed in greater detail below.

Chapter 2

In Chapter 2, we examined ethanol-responsive modules (or small, parsed out networks) in dopaminergic reward pathway-related brain regions (Prefrontal Cortex, Nucleus Accumbens, and Ventral Tegmental Area) of acute ethanol-treated mice, in combination with Alcohol Dependence GWAS data. Previous findings have suggested that gene expression networks in the mouse brain are capable of providing insight into human AD, given significant overrepresentation of human GWAS signals in mouse-derived ethanol-regulated networks (97). Therefore, we hypothesized that the integration of mouse gene expression data (with the benefits of experimental control offered by model organisms) and human GWAS data would allow us to identify novel genes associated with AD, in the context of mechanistic frameworks. For the VTA, specifically, module scores (dependent on ethanol sensitivity of expression correlations and GWAS significance of the constituent genes) significantly predicted average p-values in the independent GWAS sample. Gene ontology overrepresentation analysis revealed network functions related to ubiquitination, Syndecan and Wnt signaling, actin-related activity, and transcription regulation (274). These

results implicated: 1) that the co-analysis of mouse gene expression and human GWAS data is an effective method for identifying novel gene networks associated with alcohol-related traits; 2) that different signals between GWAS datasets are identifying similar functional pathways; 3) that the correlation between initial ethanol sensitivity (or neural reactivity) and the risk for developing dependence may be mediated by pathways primarily in the VTA, especially via ubiquitin-related mechanisms, but also by actin-mediated synaptic remodeling in the PFC, and syndecan signaling and transcription regulation in the NAc.

This analysis examined gene expression in BXD recombinant inbred mice (a genetic panel of mice which allows for technical replication across multiple mice with identical genotypes) that had been treated with ethanol, as opposed to voluntarily exposing themselves. This choice was made to capitalize on the experimental advantages of using mouse models: 1) experimental control over the environment and ethanol exposure in amount and timing; 2) technical replication; and 3) precise timing between last ethanol exposure and time of death; 4) specifically the collection of data that is expected to vary over these variables (time, environments, and timing of death) with experimental control over them. We also capitalized on the advantages of human genetic data by testing for effects of a genetic variable that is not expected to vary across these covariates (i.e. genotypes) to test their effects on behavioral measures, which are more complex than what can be exhibited by mice, with greater genomic precision and on a larger number of genetic variants than most mouse models allow (due to small numbers of founder strains and large linkage disequilibrium blocks). After performing an analysis that took full advantage of the benefits of using data from model organisms and humans, we wanted to examine related phenotypes in a mouse model that possessed features of both. We therefore elected to study voluntary ethanol consumption in Diversity Outbred stock, a

genetically (and therefore behaviorally) complex population of mice that allows for precise genetic mapping of behavioral traits across 45 million genetic variants.

Chapter 3

In Chapter 3, we determined the genetic mappability of voluntary ethanol consumption under intermittent and chronic access models, by examining the heritability of drinking behaviors in the 8 progenitor strains of the DO mouse stock. Ethanol drinking behaviors had not yet been examined in several of these strains, so this experiment had the added contribution of being the first to characterize voluntary ethanol consumption and preference in these strains. Because voluntary ethanol consumption and preference over water have been shown to differ significantly between lab-derived inbred mouse strains (implicating a genetic component of these behaviors), we anticipated that these strains (three of which are wild-derived) would yield high heritability coefficients for these phenotypes, indicating that they were genetically driven and therefore likely mappable (79). These behaviors were, in fact, highly significantly heritable across these strains, with several combinations of significant pair-wise differences between strains, indicating that each strain was contributing alleles relevant to drinking behavior. Results suggested that these differences may be driven by pharmacokinetics, but this finding was not able to undergo rigorous statistical testing due to small sample sizes.

Because heritability coefficients were slightly higher under intermittent access than under chronic access, an intermittent access paradigm was selected to be use for genetically mapping ethanol consumption and preference in DO mice.

Chapter 4

In Chapter 4, we discussed our genetic analysis of ethanol consumption and preference under an intermittent ethanol access paradigm in a sample of over 600 Diversity Outbred mice. Given that these mice have been used in recently published studies to map behavioral loci with greater precision than prior studies utilizing other mouse strains, we hypothesized that we would have the same success with respect to drinking behaviors (35). As expected, we identified several behavioral Quantitative Trait Loci for total ethanol consumption, total ethanol preference, and preference for 30% ethanol over 15% ethanol (i.e. "30% ethanol choice"), with unprecedented precision. Several candidate genes appeared in the literature as being associated with other neurological, psychiatric, and behavioral traits. These results supported findings from Sanchez-Roige et al. (25), which showed that problematic drinking is genetically correlated with other psychiatric disorders.

This experiment had the added benefit of access to longitudinal drinking data, allowing us to identify QTL that were unique to initial and long-term drinking. Because intermittent ethanol access has been shown to increase voluntary drinking over time, which is believed to be attributable to relapse-like drinking following forced abstinence, we believed that some of our identified QTL would differ between time points across ethanol exposure, reflecting the unique mechanisms involved in long-term vs. initial drinking behaviors (37). However, we also anticipated some genetic effects to be stable across time. To test this hypothesis, we mapped QTL for initial consumption by averaging drinking values over the first week of ethanol exposure, and long-term consumption by averaging values over the last week of exposure. To reduce within-mouse variation, and therefore increase power, we averaged values across the whole study to examine consistent effects, expecting to see lesser peaks for these specific QTL

for first and last week averages and higher peaks, more significant peaks for the whole study average.

As expected, we identified several QTL unique to either long-term or initial phenotypes. However, we only identified one QTL that was seen consistently across time points (although not significant for the first week timeframe), and one QTL uniquely significant for the whole study time point (Chr11 locus for total ethanol consumption). One QTL identified for whole-study phenotypes were also identified for the corresponding last-week phenotypes, but had a greater LOD score for the last-week time interval, indicating that the signal was truly related to long-term drinking but was being picked up by the whole study average. Candidate genes for the two loci that were significant for first week mean total ethanol consumption and preference showed associations with GABA signaling. The loci identified for last-week and whole-study phenotypes had many previous associations with behavioral, neurological, and psychiatric phenotypes (including alcohol-related traits), but very few with cognitive ability. Additionally, several loci were unique to the last week of consumption and the only locus for 30% ethanol choice was identified for this time interval. The top candidate genes for these QTL were largely involved in neurogenesis, growth and development, synaptic plasticity, axon guidance, and neuroinflammation. These findings indicate that long-term drinking under intermittent access is, in fact, driven by unique mechanisms, likely involving neuroinflammatory responses and rerouting of neuronal pathways in response to long-term ethanol exposure. We believe these mechanisms are likely those involved in the beginning stages of dependence.

Chapter 5

In Chapter 5, we examined ethanol-associated gene expression patterns and networks in the PFCs of the same sample of DO mice studied in Chapter 4. First, we examined individual genes for differential expression between the 100 highest drinking mice and controls, then examined the correlation of these genes with ethanol consumption within the 100 highest and 100 lowest drinkers. As expected, some genes to be significant in both analyses, but others to only be significantly associated with ethanol consumption in drinking mice. Investigators from a previous study that identified genes exhibiting similar behaviors (Bogenpohl et al., in preparation) proposed that the former genes were likely ethanol-regulated (meaning their expression levels change in response to ethanol exposure) and the latter genes were ethanol-regulating (meaning that their basal expression levels drive drinking behaviors). As expected, we did uncover genes fitting each of these descriptions, and genes that were significantly differentially expressed between high drinkers and controls but were not correlated with ethanol consumption amongst ethanol-exposed mice. These results indicate that there are many genes whose expression levels are sensitive to ethanol in a relatively dose-independent fashion, in that the ceiling of effects on expression is reached at very low doses of ethanol. These could be genes that are either very tightly regulated by feedback loops, such that only a small amount of variation is tolerated before efficient compensatory mechanisms are triggered. Although biologically less likely, they could also potentially be genes that are on the opposite end of the spectrum, and are very loosely regulated, such that any amount of perturbation yields similar, relatively unrestricted effects. Functional overrepresentation analysis revealed that genes that we believe to be ethanol-regulating were largely involved in protein- and RNA-regulatory mechanisms and immune response. This result suggests that basal differences in protein level and activity regulatory mechanisms, as well as immune-related pathways, mediate differences in

drinking behaviors. The ethanol-regulated genes, however, were overrepresented with functions related to transmembrane transport, neuron projection, and cell-maintenance (specifically, cellular respiration and metabolism, and energy use and production). Interestingly, these genes were also associated with Parkinson's, Alzheimer's, and Huntington's Diseases, all of which are neurodegenerative in nature. This is consistent with prior studies that have associated long-term alcoholism with neuronal death (306). These results suggest that, in addition to neurotransmission, ethanol affects mechanisms involved in maintaining neurons' metabolic processes and their ability to transmit signals to other neurons, likely activating synaptic plasticity and neuronal death.

To gain more refined insight into the mechanistic networks in which these genes were functioning, we mapped co-expression networks, and determined their relationship with ethanol consumption and their overrepresentation with genes that presented differential expression between high drinkers and controls. We identified multiple networks whose expression levels appeared to be either ethanol-regulating or ethanol-regulated, with respect to expression correlations with drinking behaviors and overrepresentation with genes differentially expressed between high drinkers and controls. We also discovered several networks whose eQTL overlapped with bQTL for ethanol consumption and preference, and networks that were significantly overrepresented with genes associated with drinking levels in the GSCAN GWAS sample, implicating that these genetic associations with drinking behaviors are mediated by functions in which these networks are involved. Functional overrepresentation analysis revealed that many of these networks were associated with either ubiquitin-activity, or neurite growth and remodeling. From these findings, we concluded that gene expression studies in voluntarily ethanol drinking mice are capable of providing insight to neuronal mechanisms related to

drinking behaviors in humans, and that neuroplasticity may play a role in earlier stages of dependence than previously thought. We elaborate further on this point below.

Overarching Themes and Future Directions

Consumption vs. Dependence

Twin samples have revealed significant genetic correlations between alcohol consumption and dependence (meaning they are driven by many of the same genes) (82). However, a recent GWAS study tested the ability of polygenic risk scores (or genetic risk scores based on the number of trait-related minor alleles an individual's genome possesses, weighted by their effect size) derived from alcohol consumption values in a non-dependent population to predict problematic alcohol use in another sample (25).

Results indicated that problematic alcohol use is more highly genetically correlated with other psychiatric conditions than it is with alcohol consumption. Because it is difficult to obtain deep phenotypic data from extremely large numbers of participants (largely due to limited resources and increased time commitment required from participants), many of the more recent well-powered, alcohol-related GWASs have examined drinking behaviors with the intention of identifying dependence-relevant genes. The results from Sanchez-Roige et al. (25) called into question the interpretability of results from such studies with respect to alcoholism.

Intermittent Ethanol Access voluntary consumption models have been shown to model relapse-like increases in ethanol consumption over time in certain mouse strains, and has been used to gain insight into genetic and biological mechanisms relevant to alcohol dependence (31, 36, 37, 79, 97). For

these paradigms, mice have intermittent access to alcohol with continuous access to water, so that all ethanol consumption is voluntary. Our study characterized ethanol consumption and preference for ethanol over water, under an IEA paradigm, in the eight Diversity Outbred strains, and in alignment with findings from studies performed on other strains, revealed high heritability of these traits (Chap. 3) (37) (36) (79). We then sought to identify specific genetic loci and mechanistic networks contributing to these phenotypes by mapping behavioral quantitative trait loci and ethanol-correlated gene expression networks in a large sample of DO mice. These analyses yielded results that overlapped with human studies on AD and other psychiatric disorders, as well as those on alcohol drinking, indicating that this model is capable of identifying genetic loci networks relevant to both alcohol consumption and dependence.

Specifically, genes that fell within our QTL support intervals uniquely for long-term ethanol consumption and preference had significant associations with AD in GWASs, but not with consumption or alcohol sensitivity (170). However, QTL identified for the first week and whole study possessed genes that were associated with initial alcohol sensitivity and alcohol consumption in human GWASs. We interpreted these results to mean that initial drinking values were successfully modeling drinking behaviors related to pre-dependence drinking behaviors. These behaviors are thought to be moderated by initial sensitivity and positively reinforcing motivators that drive individuals to drink before developing dependence, such as reward and anxiolysis (249). Furthermore, several of the top candidate genes (*Atp2b2*, *Slc6a1*, *Slc6a11*, and *Gphn*) in initial drinking behavior QTL were directly involved in GABA (gamma-aminbutyric acid) signaling, which is moderated by alcohol and plays a multifaceted role in the development of dependence (45)

(50) (13). These results also suggest that drinking behaviors later in the study, after enduring several alternating periods of exposure and abstinence, successfully modeled drinking behaviors in humans that are specifically associated with dependence. Although no animal model perfectly reflects human behavior, this study suggests that voluntary ethanol consumption paradigms in mice are capable of revealing biological mechanisms relevant to dependence.

In addition to the insights our studies have provided regarding animal models, our data can be interpreted in the context of shedding light on the relevance of human alcohol consumption genetic findings to alcohol dependence. Our study identified genes unique to early and long-term drinking, but also identified one locus that appeared to represents ethanol consumption relatively stably across time (on Chr1 for total ethanol preference). This locus replicated results from the recent GSCAN GWAS study on alcohol consumption (61). The most empirically supported gene in this region was *Nucks1*, which has relatively unknown functionality, but has been found to be associated with neurodegenerative diseases and inflammation (184) (185) (186) (183). Given that this QTL was most highly significant for the whole study time interval than it was for either the first (not significant, but a trending peak) or last week timeframes, we believe that this gene may be specifically involved in neuronal mechanisms that drive non- dependent (or not yet dependent) individuals to consume alcohol. Given that this locus appeared in both humans and singly-house mice, these mechanisms are likely related to psychotropic reward, as opposed to qualities that might be unique to social drinking, such as personality traits or social anxiety (8).

Future studies should examine the effects of mutations in the gene *Nucks1* on subjective reward, and neurological pathways involved in reward, such as those involved in the mesolimbocortical. This could be done by directly mutating the gene, or changing its expression levels and examining its effects on drinking behavior in a model organism. These studies should specifically focus on drinking behaviors after consumption levels have become stable over time, as our results indicate that this gene may be associated with initial drinking levels, but its effects are more evident on a longer term scale. Such studies should also examine specifically more negative-reinforcement-driven behaviors in these mice, such as craving- or withdrawal-driven self-administration. These results could determine whether *Nucks1* is related to the negative reinforcing properties of alcohol seen in the later stages of dependence development, or if it is unique to positive-reinforcement-driven substance use, as our findings suggest. Finally, the relationship between this gene and neuroinflammation and neuronal death should be explored. It is known that alcoholism is associated with decreased white matter and increased inflammation in the brain (264, 265). Because this gene appears to be associated with positive reward-driven drinking, these studies may reveal interactions between alcohol- induced overstimulation of dopaminergic neurons and neurotoxicity.

Ubiquitin and Alcohol

Both the human-mouse integrative gene expression network study on acute ethanol-treated BXDs (Chap. 2), and the study on gene expression in the brain of voluntarily drinking Diversity Outbred mice (Chap. 5), several networks were associated with regulation of protein levels, at transcriptional, translational, and post-translational levels. In particular, ubiquitin activity was overrepresented in: the discussed BXD Ventral Tegmental Area network (Bisque);

the ethanol-regulating genes in the DO PFC; and DO PFC modules Turquoise and Light Green. Showing significant overrepresentation with differentially expressed genes between DO high drinkers and controls, Turquoise was considered to be an ethanol-regulated module, whereas Blue and Light Green were not. Pairing this information with the knowledge that similar functional overrepresentation appeared in our acute-ethanol-treated sample, these results collectively suggest that these protein-level regulatory activities begin responding to alcohol upon initial exposure and continue to do so through long-term exposure. Furthermore, these ubiquitin-related modules were identified in two distinct brain regions involved in the dopaminergic reward pathway (VTA and PFC).

Multiple other studies have identified significant associations between ethanol-related behaviors and ubiquitin genes (307 & Ebihara, 2013, 308 2014). One study in particular, conducted by Melendez et al. (309 & Becker) identified genes with altered expression levels in the PFC (as well as other brain regions) related to involuntary chronic intermittent ethanol exposure (via ethanol vapor), and performed functional overrepresentation analysis on significant genes. Taking a more targeted approach, one study has specifically examined the response of ethanol-related phenotypes to the down regulation of a specific ubiquitin gene. Mice exhibited a decreased ethanol preference and decreased recovery time from the sedating effects of ethanol, in response to systematic down regulation of ubiquitin-specific peptidase 46, which is involved in deubiquitination (although the exact proteins with which it interacts are unknown) (307). Additionally, this particular protein has been found to be ubiquitously expressed in the frontal lobe of mice (310). These results suggest that the ubiquitin proteins with which this enzyme interacts are partly responsible for increased ethanol sensitivity of these mice,

possibly specifically in the frontal lobe. Although ubiquitin proteins are involved in many different processes, due to their interactions with a vast array of proteins, this example provides further evidence that ubiquitin-related mechanisms do moderate ethanol-related behaviors, and even specifically ethanol preference.

Other study have shown both increases and decreases in expression levels of ubiquitin genes in brain white matter and cortical neurons in alcoholics compared to controls (308) (308) (311) (312).

Furthermore, ubiquitin activity has been shown to be sensitive to changes in neuronal signaling, resulting in the ubiquitination of synaptic and cytoskeletal proteins, implicating its involvement in neuroplasticity (313). Ubiquitin has also been implicated in the process of developing long-term potentiation, which also suggests involvement in synaptic plasticity (314). As discussed in more detail below, several of our analyses identified genes related to actin activity, extracellular matrix integrity, and other processes that suggest a role of synaptic remodeling, not just in the later stages of dependence, but in the beginning stages, as well. We believe that these changes are likely ubiquitin-mediated. Future studies should examine the effects of down- and up-regulation of ubiquitin-related genes identified in our modules (ex.

Trip12) in a similar fashion, examining differences in voluntary ethanol consumption and preference between these mice and wild type controls, but also determining the effects of ethanol on these exact systems after ethanol treatment or exposure. They should also study the effects of these genes on other genes specifically in our ubiquitin-overrepresented modules, particularly those related to the neurite outgrowth and routing involved in neuroplastic processes. Although ubiquitin genes are not likely to be feasible pharmacotherapeutic targets for alcoholism, due to their roles in cell maintenance and stability, identifying proteins targeted by these genes could be

illuminating with respect to the exact pathways involved in alcohol-mediated neuroplasticity, and therefore the development of addiction.

Neuroplasticity in Early Dependence

Some potential dependence-relevant targets of ubiquitin could be proteins involved specifically in neurite outgrowth and axon guidance. Our studies identified several behavioral QTL, and ethanol-associated networks in the PFC in both BXD and DO mice that were associated with cell structural and scaffolding proteins. Specifically, the bQTL that were specific to long-term (i.e. last week) drinking behaviors in the DO mice contained genes that were largely involved in neurite growth, actin activity, and extracellular matrix integrity. A gene that contained all of the top-scoring SNPs on the Chr11 QTL for last week mean total ethanol consumption in the DO mice (*Adamts2*) acts as a protease that moderates neuroplastic changes by acting on proteins in the extracellular matrix. The module Alice Blue in the PFC of acute-ethanol exposed BXD mice, which contained a significant number of AD-associated genes spanning across two independent GWAS datasets, was significantly overrepresented with functions related to actin. Modules Blue, Dark Turquoise, Pale Turquoise, and Light Green in the DO mouse PFC were also overrepresented with functions related to actin-mediated neurogenesis, neurite growth, and axonal re-routing. The Blue module was identified as being "ethanol regulating", indicating that basal expression levels or sensitivity of these proteins drives drinking behaviors. Furthermore, both this module and Light Green were overrepresented with genes associated with alcohol consumption in the GSCAN GWAS. Finally, three of these four modules' Eigen Gene eQTLs overlapped with the *Nucks1*-containing bQTL for last week total ethanol consumption on Chr1 discussed above; and the remaining module's (Pale Turquoise)

Eigen Gene eQTL overlapped with the Chr16 bQTL for last week mean preference, which contained top candidate genes with functions related to neurogenesis, axon guidance, and synaptic plasticity. Together, these results implicate that genes related to neurite outgrowth and axon guidance moderate drinking behaviors. We believe that the sensitivity of these genes to ethanol varies across alleles, which moderates the extent to which neuroplastic changes occur in response to ethanol, and thereby affect the risk for developing alcoholism.

Addiction is thought to be a lifelong disease, because of irreversible changes that occur in the brain in response to chronic drug exposure resulting in over- and under-stimulation of primarily, but certainly not only, glutamatergic, GABAergic, dopaminergic, and serotonergic pathways in the brain. However, these changes are thought to occur during the late-stages of dependence development, when signs of withdrawal and craving begin to be presented (9, 249). All of these modules and QTL have been identified in acutely ethanol-exposed mice or mice exhibiting increased levels of ethanol consumption over time, modeling the very early stages of dependence development.

Furthermore, at least one of the relevant modules (DO PFC Blue) is thought to be ethanol-regulating based on our results, and other modules were overrepresented with GWAS signals for dependence or alcohol consumption, indicating that basal genetic differences drive drinking behaviors. Neuroplastic changes related to alcohol dependence are thought to occur in response to the weakening and strengthening of synaptic signals that are caused by gene expression changes moderated by long-term exposure to alcohol (9, 249). However, our results indicate that mechanisms related to synaptic remodeling and axon guidance are relevant much earlier on in the process of dependence development.

Future studies should examine the effects of acute and short-term alcohol exposure on proteins involved in changes in cytoskeletal proteins (particularly actin) and extra-cellular scaffolding proteins in and around neurons. These pathways could be associated with the rewarding properties of alcohol that lead to abuse, as our results indicate that they may not be involved not only in the transition from abusive to addictive use, but also in early drinking behaviors and abuse tendencies. We hypothesize that these studies will find that the innate sensitivity of cytoskeletal and scaffolding proteins affects an individual's propensity for drinking alcohol abusively, and therefore for developing dependence.

Final Conclusions

This collection of studies examined the genetics of alcohol consumption and dependence by analyzing gene expression networks and genetic associations in genetically complex mouse strains, in the context of ethanol treatment and voluntary ethanol consumption. By integrating this data with human alcohol dependence and drinking genetic association data, we were able to identify several mechanistic pathways relevant to these two phenotypes. Specifically, our results suggested involvement of functions related to ubiquitination and actin-mediated neurite growth and axon guidance in drinking behaviors that occur before and during the early stages of the development of dependence. They also implicated that alcohol consumption models in mice are capable of modeling both non-pathological early drinking behaviors, as well as abusive drinking that leads to the development of addiction. Finally, our study on voluntary ethanol drinking in a genetically diverse mouse stock replicated results from a human genome wide association study on alcohol consumption, by identifying a locus on Chr1 relevant to long-term consumption, implicating the involvement in the gene *Nucks1* in drinking behaviors. Overall, these studies

indicated that negative protein regulation and synaptic plasticity moderate drinking behaviors exhibited during the early, reward-driven stages of the development of alcohol dependence.

References

1. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2010;380(9859):2224-60.
2. Witt SH, Streit F, Jungkunz M, Frank J, Awasthi S, Reinbold CS, et al. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. Transl Psychiatry. 2017;7(6):e1155.
3. Network GBoDC. Global Burden of Disease Study 2016 (GBD 2016) Alcohol Use Estimates 1990-2016. In: National Institue on Aging NIoH, editor. Seattle, Washington, United States: Institution for Health Metris and Evaluation (IHME), 2018; 2018.
4. Takahashi T, Lapham G, Chavez LJ, Lee AK, Williams EC, Richards JE, et al. Comparison of DSM-IV and DSM-5 criteria for alcohol use disorders in VA primary care patients with frequent heavy drinking enrolled in a trial. Addict Sci Clin Pract. 2017;12(1):17.
5. Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, et al. Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. JAMA Psychiatry. 2015;72(8):757-66.
6. Deborah S. Hasin PD, Charles P. O'Brien, M.D., Ph.D., Marc Auriacombe, M.D., Guilherme Borges, Sc.D., Kathleen Bucholz, Ph.D., Alan Budney, Ph.D., Wilson M. Compton, M.D., M.P.E., Thomas Crowley, M.D., Walter Ling, M.D., Nancy M. Petry, Ph.D., Marc Schuckit, M.D., and Bridget F. Grant, Ph.D. DSM-5 Criteria for Substance Use Disorders: Recommendations and Rationale. American Journal of Psychiatry. 2013;170(8):834-51.
7. Alcohol Use Disorder: A Comparison Between DSM-IV and DSM-V. In: Alcoholism NIoAAa, editor. 2016.
8. Koob GF, Volkow ND. Neurocircuitry of Addiction. Neuropsychopharmacology. 2009;35:217.
9. Miles MF. Alcohol's Effect on Gene Expression. Alcohol Health and Research World. 1995;19(3):237-43.
10. Ballenger JC, Post RM. Kindling as a model for alcohol withdrawal syndromes. Br J Psychiatry. 1978;133:1-14.
11. Moos RH, Moos BS. Rates and predictors of relapse after natural and treated remission from alcohol use disorders. Addiction. 2006;101(2):212-22.
12. Antonelli M, Ferrulli A, Sestito L, Vassallo GA, Tarli C, Mosoni C, et al. Alcohol addiction - the safety of available approved treatment options. Expert opinion on drug safety. 2018;17(2):169-77.
13. Agrawal A, Bierut LJ. Identifying genetic variation for alcohol dependence. Alcohol Res. 2012;34(3):274-81.
14. Handbook of Clinical Neurology: Alcohol and the Nervous System. 3 ed. Michael J. Aminoff FB, and Dick Swaab, editor. Edenborough, London, New York, Oxford, Philadelphia, St. Louis, Sydney, Toronto: Elsevier; 2014.

15. Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. Lancet Psychiatry. 2016;3(8):760-73.
16. Oscar-Berman M, Marinković K. Alcohol: effects on neurobehavioral functions and the brain. Neuropsychology review. 2007;17(3):239-57.
17. Bogg T, Fukunaga R, Finn PR, Brown JW. Cognitive control links alcohol use, trait disinhibition, and reduced cognitive capacity: Evidence for medial prefrontal cortex dysregulation during reward-seeking behavior. Drug Alcohol Depend. 2012;122(1-2):112-8.
18. Pfefferbaum A, Rosenbloom M, Serventi KL, Sullivan EV. Corpus callosum, pons, and cortical white matter in alcoholic women. Alcohol Clin Exp Res. 2002;26(3):400-6.
19. Mons N, Beracochea D. Behavioral Neuroadaptation to Alcohol: From Glucocorticoids to Histone Acetylation. Front Psychiatry. 2016;7:165.
20. Cotton NS. The familial incidence of alcoholism: a review. J Stud Alcohol. 1979;40(1):89-116.
21. Verhulst B, Neale MC, Kendler KS. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. Psychol Med. 2015;45(5):1061-72.
22. Prescott CA, Kendler KS. Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. Am J Psychiatry. 1999;156(1):34-40.
23. Hart AB, Kranzler HR. Alcohol Dependence Genetics: Lessons Learned From Genome-Wide Association Studies (GWAS) and Post-GWAS Analyses. Alcohol Clin Exp Res. 2015;39(8):1312-27.
24. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.
25. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, Adams MJ, Howard DM, et al. Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. Am J Psychiatry. 2019;176(2):107-18.
26. Clarke TK, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). Mol Psychiatry. 2017;22(10):1376-84.
27. Wolen AR, C. A. Phillips, M. A. Langston, A. H. Putman, P. J. Vorster, N. A. Bruce, T. P. York, R. W. Genetic dissection of acute ethanol responsive gene networks in prefrontal cortex: functional and mechanistic implications. PLoS One. 2012;7(4):e33575.
28. Smith ML, Lopez MF, Archer KJ, Wolen AR, Becker HC, Miles MF. Time-Course Analysis of Brain Regional Expression Network Responses to Chronic Intermittent Ethanol and Withdrawal: Implications for Mechanisms Underlying Excessive Ethanol Consumption. PLoS One. 2016;11(1):e0146257.
29. Mamdani M, Williamson V, McMichael GO, Blevins T, Aliev F, Adkins A, et al. Integrating mRNA and miRNA Weighted Gene Co-Expression Networks with eQTLs in the Nucleus Accumbens of Subjects with Alcohol Dependence. PLoS One. 2015;10(9):e0137671.
30. Farris SP, Miles MF. Fyn-dependent gene networks in acute ethanol sensitivity. PLoS One. 2013;8(11):e82435.

31. Wolstenholme JT, Warner JA, Capparuccini MI, Archer KJ, Shelton KL, Miles MF. Genomic Analysis of Individual Differences in Ethanol Drinking: Evidence for Non-Genetic Factors in C57BL/6 Mice. PLoS ONE. 2011;6(6):e21100.
32. Chesler EJ. Out of the bottleneck: the Diversity Outcross and Collaborative Cross mouse populations in behavioral genetics research. Mamm Genome. 2014;25(1-2):3-11.
33. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52-8.
34. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88.
35. Logan RW, Robledo RF, Recla JM, Philip VM, Bubier JA, Jay JJ, et al. High-precision genetic mapping of behavioral traits in the diversity outbred mouse population. Genes Brain Behav. 2013;12(4):424-37.
36. Hwa LS, Chu A, Levinson SA, Kayyali TM, DeBold JF, Miczek KA. Persistent escalation of alcohol drinking in C57BL/6J mice with intermittent access to 20% ethanol. Alcohol Clin Exp Res. 2011;35(11):1938-47.
37. Melendez RI. Intermittent (every-other-day) drinking induces rapid escalation of ethanol intake and preference in adolescent and adult C57BL/6J mice. Alcohol Clin Exp Res. 2011;35(4):652-8.
38. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
39. Felson J. What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. Soc Sci Res. 2014;43:184-99.
40. Mbarek H, Milaneschi Y, Fedko IO, Hottenga JJ, de Moor MH, Jansen R, et al. The genetics of alcohol dependence: Twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. Am J Med Genet B Neuropsychiatr Genet. 2015;168(8):739-48.
41. Edenberg HJ, Foroud T, Koller DL, Goate A, Rice J, Van Eerdewegh P, et al. A family-based analysis of the association of the dopamine D2 receptor (DRD2) with alcoholism. Alcohol Clin Exp Res. 1998;22(2):505-12.
42. Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, et al. Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet. 1998;81(3):207-15.
43. Tanna VL, Wilson AF, Winokur G, Elston RC. Possible linkage between alcoholism and esterase-D. J Stud Alcohol. 1988;49(5):472-6.
44. Hill SY, Zezza N, Wipprecht G, Xu J, Neiswanger K. Linkage studies of D2 and D4 receptor genes and alcoholism. Am J Med Genet. 1999;88(6):676-85.
45. Edenberg HJ, Foroud T. Genetics and alcoholism. Nat Rev Gastroenterol Hepatol. 2013;10(8):487-94.
46. Deak JD, Miller AP, Gizer IR. Genetics of alcohol use disorder: a review. Curr Opin Psychol. 2018;27:56-61.
47. Dick DM, Foroud T. Candidate genes for alcohol dependence: a review of genetic evidence from human studies. Alcohol Clin Exp Res. 2003;27(5):868-79.
48. 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;11(6):603-11.
49. Ducci F, Goldman D. Genetic approaches to addiction: genes and alcohol. Addiction. 2008;103(9):1414-28.
50. Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, et al. Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. Am J Hum Genet. 2004;74(4):705-14.
51. Long JC, Knowler WC, Hanson RL, Robin RW, Urbanek M, Moore E, et al. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an American Indian population. Am J Med Genet. 1998;81(3):216-21.
52. Goldman D. The missing heritability of behavior: the search continues. Psychophysiology. 2014;51(12):1327-8.
53. Sullivan PF. Spurious genetic associations. Biol Psychiatry. 2007;61(10):1121-6.
54. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-27.
55. Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. Science. 2015;349(6255):1489-94.
56. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest. 2008;118(5):1590-605.
57. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320.
58. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008;32(4):381-5.
59. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461.
60. Schumann G, Liu C, O'Reilly P, Gao H, Song P, Xu B, et al. KLB is associated with alcohol drinking, and its gene product beta-Klotho is necessary for FGF21 regulation of alcohol preference. Proc Natl Acad Sci U S A. 2016;113(50):14372-7.
61. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nature genetics. 2019;51(2):237-44.
62. Foroud T, Li T-K. Genetics of Alcoholism: A Review of Recent Studies in Human and Animal Models. The American Journal on Addictions. 1999;8(4):261-78.
63. Mayfield J, Arends MA, Harris RA, Blednov YA. Genes and Alcohol Consumption: Studies with Mutant Mice. Int Rev Neurobiol. 2016;126:293-355.
64. Silver L. Mouse Genetics. new York, NY, United States: Oxford University Press; 1995.
65. Belknap JK, Atkins AL. The replicability of QTLs for murine alcohol preference drinking behavior across eight independent studies. Mamm Genome. 2001;12(12):893-9.
66. Cockram J, Mackay I. Genetic Mapping Populations for Conducting High-Resolution Trait Mapping in Plants. Adv Biochem Eng Biotechnol. 2018;164:109-38.
67. Tolliver BK, Belknap JK, Woods WE, Carney JM. Genetic analysis of sensitization and tolerance to cocaine. J Pharmacol Exp Ther. 1994;270(3):1230-8.

68. Grisel JE, Belknap JK, O'Toole LA, Helms ML, Wenger CD, Crabbe JC. Quantitative trait loci affecting methamphetamine responses in BXD recombinant inbred mouse strains. J Neurosci. 1997;17(2):745-54.
69. Putman AH, Wolen AR, Harenza J, Yordanova RK, Webb BT, Chesler EJ, et al. Identification of Quantitative Trait Loci and Candidate Genes for an Anxiolytic-like Response to Ethanol in BXD Recombinant Inbred Strains. Genes Brain Behav. 2016.
70. Browman KE, Crabbe JC. Quantitative trait loci affecting ethanol sensitivity in BXD recombinant inbred mice. Alcohol Clin Exp Res. 2000;24(1):17-23.
71. Risinger FO, Cunningham CL. Ethanol-induced conditioned taste aversion in BXD recombinant inbred mice. Alcohol Clin Exp Res. 1998;22(6):1234-44.
72. Taylor BA, Wnek C, Kotlus BS, Roemer N, MacTaggart T, Phillips SJ. Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps. Mamm Genome. 1999;10(4):335-48.
73. Churchill GA. Reinventing the laboratory mouse. National Institute of Environmental Health Sciences; 2014.
74. Philip VM, Sokoloff G, Ackert-Bicknell CL, Striz M, Branstetter L, Beckmann MA, et al. Genetic analysis in the Collaborative Cross breeding population. Genome Res. 2011;21(8):1223-38.
75. Woods LC, Mott R. Heterogeneous Stock Populations for Analysis of Complex Traits. Methods Mol Biol. 2017;1488:31-44.
76. Churchill GA, Gatti DM, Munger SC, Svenson KL. The Diversity Outbred mouse population. Mamm Genome. 2012;23(9-10):713-8.
77. Crabbe JC. Use of animal models of alcohol-related behavior. Handb Clin Neurol. 2014;125:71-86.
78. Yoneyama N, Crabbe JC, Ford MM, Murillo A, Finn DA. Voluntary ethanol consumption in 22 inbred mouse strains. Alcohol. 2008;42(3):149-60.
79. Rosenwasser AM, Fixaris MC, Crabbe JC, Brooks PC, Ascheid S. Escalation of intake under intermittent ethanol access in diverse mouse genotypes. Addict Biol. 2013;18(3):496-507.
80. Lopez MF, Becker HC. Effect of pattern and number of chronic ethanol exposures on subsequent voluntary ethanol intake in C57BL/6J mice. Psychopharmacology (Berl). 2005;181(4):688-96.
81. Spanagel R. Alcohol addiction research: from animal models to clinics. Best Pract Res Clin Gastroenterol. 2003;17(4):507-18.
82. Grant JD, Agrawal A, Bucholz KK, Madden PA, Pergadia ML, Nelson EC, et al. Alcohol consumption indices of genetic risk for alcohol dependence. Biol Psychiatry. 2009;66(8):795-800.
83. Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. 1955;7(3):277-318.
84. Manichaikul A, Dupuis J, Sen S, Broman KW. Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics. 2006;174(1):481-9.
85. Melo JA, Shendure J, Pociask K, Silver LM. Identification of sex–specific quantitative trait loci controlling alcohol preference in C57BL/6 mice. Nature genetics. 1996;13:147.

86. Erwin VG, Markel PD, Johnson TE, Gehle VM, Jones BC. Common quantitative trait loci for alcohol-related behaviors and central nervous system neurotensin measures: hypnotic and hypothermic effects. J Pharmacol Exp Ther. 1997;280(2):911-8.
87. Saba LM, Bennett B, Hoffman PL, Barcomb K, Ishii T, Kechris K, et al. A systems genetic analysis of alcohol drinking by mice, rats and men: influence of brain GABAergic transmission. Neuropharmacology. 2011;60(7-8):1269-80.
88. Hoffman PL, Saba LM, Vanderlinden LA, Tabakoff B. Voluntary exposure to a toxin: the genetic influence on ethanol consumption. Mamm Genome. 2018;29(1-2):128-40.
89. Mulligan MK, Zhao W, Dickerson M, Arends D, Prins P, Cavigelli SA, et al. Genetic Contribution to Initial and Progressive Alcohol Intake Among Recombinant Inbred Strains of Mice. Front Genet. 2018;9:370.
90. Koide T, Ikeda K, Ogasawara M, Shiroishi T, Moriwaki K, Takahashi A. A new twist on behavioral genetics by incorporating wild-derived mouse strains. Exp Anim. 2011;60(4):347-54.
91. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature. 2007;445(7124):168-76.
92. Farris SP, Arasappan D, Hunicke-Smith S, Harris RA, Mayfield RD. Transcriptome organization for chronic alcohol abuse in human brain. Mol Psychiatry. 2015;20(11):1438-47.
93. Liu J, Lewohl JM, Harris RA, Iyer VR, Dodd PR, Randall PK, et al. Patterns of Gene Expression in the Frontal Cortex Discriminate Alcoholic from Nonalcoholic Individuals. Neuropsychopharmacology. 2005;31(7):1574-82.
94. Lewohl JM, Wang L, Miles MF, Zhang L, Dodd PR, Harris RA. Gene expression in human alcoholism: microarray analysis of frontal cortex. Alcohol Clin Exp Res. 2000;24(12):1873-82.
95. Osterndorff-Kahanek E, Ponomarev I, Blednov YA, Harris RA. Gene expression in brain and liver produced by three different regimens of alcohol consumption in mice: comparison with immune activation. PLoS One. 2013;8(3):e59870.
96. Kerns RT, Ravindranathan A, Hassan S, Cage MP, York T, Sikela JM, et al. Ethanol-responsive brain region expression networks: implications for behavioral responses to acute ethanol in DBA/2J versus C57BL/6J mice. J Neurosci. 2005;25(9):2255-66.
97. van der Vaart AD, Wolstenholme JT, Smith ML, Harris GM, Lopez MF, Wolen AR, et al. The allostatic impact of chronic ethanol on gene expression: A genetic analysis of chronic intermittent ethanol treatment in the BXD cohort. Alcohol. 2017;58:93-106.
98. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004;5(2):101-13.
99. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. Cell. 2011;144(6):986-98.
100. Horvath S. Weighted Network Analysis: Applications in Genomics and Systems Biology. New York, NY, United States: Springer; 2011.
101. Zhang H, Wang F, Xu H, Liu Y, Liu J, Zhao H, et al. Differentially co-expressed genes in postmortem prefrontal cortex of individuals with alcohol use disorders: influence on alcohol metabolism-related pathways. Hum Genet. 2014;133(11):1383-94.

102. Callaway DS, Newman ME, Strogatz SH, Watts DJ. Network robustness and fragility: percolation on random graphs. Phys Rev Lett. 2000;85(25):5468-71.
103. Han S, Yang BZ, Kranzler HR, Liu X, Zhao H, Farrer LA, et al. Integrating GWASs and human protein interaction networks identifies a gene subnetwork underlying alcohol dependence. Am J Hum Genet. 2013;93(6):1027-34.
104. Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. Nucleic acids research. 2018;46(D1):D836-d42.
105. Dowell RD. The similarity of gene expression between human and mouse tissues. Genome Biol. 2011;12(1):101.
106. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci. 2015;18(2):199-209.
107. Mamdani M, Williamson V, McMichael GO, Blevins T, Aliev F, Adkins A, et al. Integrating mRNA and miRNA Weighted Gene Co-Expression Networks with eQTLs in the Nucleus Accumbens of Subjects with Alcohol Dependence. PloS one. 2015;10(9):e0137671.
108. Network, Pathway Analysis Subgroup of Psychiatric Genomics C. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nature neuroscience. 2015;18(2):199-209.
109. Adkins AE, Hack LM, Bigdeli TB, Williamson VS, McMichael GO, Mamdani M, et al. Genomewide Association Study of Alcohol Dependence Identifies Risk Loci Altering Ethanol-Response Behaviors in Model Organisms. Alcoholism, clinical and experimental research. 2017;41(5):911-28.
110. Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. Bioinformatics (Oxford, England). 2011;27(1):95-102.
111. Wang Q, Yu H, Zhao Z, Jia P. EW_dmGWAS: edge-weighted dense module search for genome-wide association studies and gene expression profiles. Bioinformatics (Oxford, England). 2015;31(15):2591-4.
112. de Wit H, Phillips TJ. Do initial responses to drugs predict future use or abuse? Neurosci Biobehav Rev. 2012;36(6):1565-76.
113. Trim RS, Schuckit MA, Smith TL. The relationships of the level of response to alcohol and additional characteristics to alcohol use disorders across adulthood: a discrete-time survival analysis. Alcohol Clin Exp Res. 2009;33(9):1562-70.
114. Buck KJ, Metten P, Belknap JK, Crabbe JC. Quantitative trait loci involved in genetic predisposition to acute alcohol withdrawal in mice. J Neurosci. 1997;17(10):3946-55.
115. Putman AH, Wolen AR, Harenza JL, Yordanova RK, Webb BT, Chesler EJ, et al. Identification of quantitative trait loci and candidate genes for an anxiolytic-like response to ethanol in BXD recombinant inbred strains. Genes, brain, and behavior. 2016;15(4):367-81.
116. Sun Y, Zhang Y, Wang F, Sun Y, Shi J, Lu L. From genetic studies to precision medicine in alcohol dependence. Behavioural pharmacology. 2016;27(2-3 Spec Issue):87-99.

117. Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, et al. Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. Mol Psychiatry. 2014;19(1):41-9.
118. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. American Journal of Human Genetics. 2010;86(4):581-91.
119. Wang L, Jia P, Wolfinger RD, Chen X, Zhao Z. Gene set analysis of genome-wide association studies: methodological issues and perspectives. Genomics. 2011;98(1):1-8.
120. Wolen AR, Phillips CA, Langston MA, Putman AH, Vorster PJ, Bruce NA, et al. Genetic dissection of acute ethanol responsive gene networks in prefrontal cortex: functional and mechanistic implications. PloS one. 2012;7(4):e33575.
121. Osterndorff-Kahanek EA, Tiwari GR, Lopez MF, Becker HC, Harris RA, Mayfield RD. Long-term ethanol exposure: Temporal pattern of microRNA expression and associated mRNA gene networks in mouse brain. PloS one. 2018;13(1):e0190841.
122. van der Vaart A, Meng X, Bowers MS, Batman AM, Aliev F, Farris SP, et al. Glycogen synthase kinase 3 beta regulates ethanol consumption and is a risk factor for alcohol dependence. Neuropsychopharmacology. 2018;43(13):2521-31.
123. Zhang L, Wang L, Ravindranathan A, Miles MF. A new algorithm for analysis of oligonucleotide arrays: application to expression profiling in mouse brain regions. J Mol Biol. 2002;317(2):225-35.
124. Kennedy RE, Archer KJ, Miles MF. Empirical validation of the S-Score algorithm in the analysis of gene expression data. BMC Bioinformatics. 2006;7:154.
125. Thornton T, McPeek MS. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. Am J Hum Genet. 2007;81(2):321-37.
126. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature genetics. 2012;44(8):955-9.
127. Li MX, Sham PC, Cherny SS, Song YQ. A knowledge-based weighting framework to boost the power of genome-wide association studies. PLoS One. 2010;5(12):e14480.
128. Edwards AC, Aliev F, Wolen AR, Salvatore JE, Gardner CO, McMahon G, et al. Genomic influences on alcohol problems in a population-based sample of young adults. Addiction. 2015;110(3):461-70.
129. Wu J, Vallenius T, Ovaska K, Westermarck J, Makela TP, Hautaniemi S. Integrated network analysis platform for protein-protein interactions. Nat Methods. 2009;6(1):75-7.
130. Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM, et al. PINA v2.0: mining interactome modules. Nucleic acids research. 2012;40(Database issue):D862-5.
131. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. Methods Mol Biol. 2007;406:89-112.
132. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7):1177-86.
133. Lu JY, Schneider RJ. Tissue distribution of AU-rich mRNA-binding proteins involved in regulation of mRNA decay. J Biol Chem. 2004;279(13):12974-9.

134. Tiruchinapalli DM, Caron MG, Keene JD. Activity-dependent expression of ELAV/Hu RBPs and neuronal mRNAs in seizure and cocaine brain. J Neurochem. 2008;107(6):1529-43.
135. Wilke N, Sganga M, Barhite S, Miles MF. Effects of alcohol on gene expression in neural cells. Exs. 1994;71:49-59.
136. Seo D, Sinha R. Neuroplasticity and Predictors of Alcohol Recovery. Alcohol Res. 2015;37(1):143-52.
137. Kyzar EJ, Pandey SC. Molecular mechanisms of synaptic remodeling in alcoholism. Neurosci Lett. 2015;601:11-9.
138. Hu HT, Hsueh YP. Calcium influx and postsynaptic proteins coordinate the dendritic filopodium-spine transition. Dev Neurobiol. 2014;74(10):1011-29.
139. Stamatakou E, Hoyos-Flight M, Salinas PC. Wnt Signalling Promotes Actin Dynamics during Axon Remodelling through the Actin-Binding Protein Eps8. PLoS One. 2015;10(8):e0134976.
140. Chen CH, He CW, Liao CP, Pan CL. A Wnt-planar polarity pathway instructs neurite branching by restricting F-actin assembly through endosomal signaling. PLoS Genet. 2017;13(4):e1006720.
141. Bjork K, Rimondini R, Hansson AC, Terasmaa A, Hyytia P, Heilig M, et al. Modulation of voluntary ethanol consumption by beta-arrestin 2. Faseb j. 2008;22(7):2552-60.
142. Heath AC, Madden PA, Bucholz KK, Dinwiddie SH, Slutske WS, Bierut LJ, et al. Genetic differences in alcohol sensitivity and the inheritance of alcoholism risk. Psychol Med. 1999;29(5):1069-81.
143. Morean ME, Corbin WR. Subjective response to alcohol: a critical review of the literature. Alcohol Clin Exp Res. 2010;34(3):385-95.
144. Ray LA, Mackillop J, Monti PM. Subjective responses to alcohol consumption as endophenotypes: advancing behavioral genetics in etiological and treatment models of alcoholism. Subst Use Misuse. 2010;45(11):1742-65.
145. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nature genetics. 2019;51(2):237-44.
146. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, andMe Research Team tSUDWGotPGC, Adams MJ, et al. Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. The American journal of psychiatry. 2019;176(2):107-18.
147. Farris SP, Miles MF. Ethanol modulation of gene networks: implications for alcoholism. Neurobiol Dis. 2012;45(1):115-21.
148. Melendez RI, Middaugh LD, Kalivas PW. Development of an alcohol deprivation and escalation effect in C57BL/6J mice. Alcohol Clin Exp Res. 2006;30(12):2017-25.
149. Porcu P, O'Buckley TK, Lopez MF, Becker HC, Miles MF, Williams RW, et al. Initial genetic dissection of serum neuroactive steroids following chronic intermittent ethanol across BXD mouse strains. Alcohol. 2017;58:107-25.
150. Lopez MF, Miles MF, Williams RW, Becker HC. Variable effects of chronic intermittent ethanol exposure on ethanol drinking in a genetically diverse mouse cohort. Alcohol. 2017;58:73-82.

151. Ottman R. Gene-environment interaction: definitions and study designs. Prev Med. 1996;25(6):764-70.
152. Loney KD, Uddin KR, Singh SM. Strain‐Specific Brain Metallothionein II (MT‐II) Gene Expression, Its Ethanol Responsiveness, and Association With Ethanol Preference in Mice. Alcoholism: Clinical and Experimental Research. 2003;27(3):388-95.
153. Bachmanov AA. Intake of Ethanol, Sodium Chloride, Sucrose, Citric Acid, and Quinine Hydrochloride Solutions by Mice: A Genetic Analysis. 1996;26(6):563-73.
154. Crabbe JC, Cotnam CJ, Cameron AJ, Schlumbohm JP, Rhodes JS, Metten P, et al. Strain differences in three measures of ethanol intoxication in mice: the screen, dowel and grip strength tests. Genes Brain Behav. 2003;2(4):201-13.
155. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, et al. High-resolution genetic mapping using the Mouse Diversity outbred population. Genetics. 2012;190(2):437-47.
156. Gatti DM, Svenson KL, Shabalin A, Wu LY, Valdar W, Simecek P, et al. Quantitative trait locus mapping methods for diversity outbred mice. G3 (Bethesda). 2014;4(9):1623-33.
157. Vanderlinden LA, Saba LM, Bennett B, Hoffman PL, Tabakoff B. Influence of sex on genetic regulation of "drinking in the dark" alcohol consumption. Mamm Genome. 2015;26(1-2):43-56.
158. Chesler EJ, Plitt A, Fisher D, Hurd B, Lederle L, Bubier JA, et al. Quantitative trait loci for sensitivity to ethanol intoxication in a C57BL/6Jx129S1/SvImJ inbred mouse cross. Mamm Genome. 2012;23(5-6):305-21.
159. Buck KJ, Rademacher BS, Metten P, Crabbe JC. Mapping murine loci for physical dependence on ethanol. Psychopharmacology (Berl). 2002;160(4):398-407.
160. Morgan AP, Fu CP, Kao CY, Welsh CE, Didion JP, Yadgary L, et al. The Mouse Universal Genotyping Array: From Substrains to Subspecies. G3 (Bethesda). 2015;6(2):263-79.
161. 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;81.
162. Reed DR.
163. 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;81(3):559-75.
164. Purcell S. PLINK. 1.9 ed2016.
165. Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen S, et al. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. Genetics. 2019;211(2):495-502.
166. Cheng R, Parker CC, Abney M, Palmer AA. Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. G3 (Bethesda). 2013;3(10):1861-7.
167. Recla JM, Robledo RF, Gatti DM, Bult CJ, Churchill GA, Chesler EJ. Precise genetic mapping and integrative bioinformatics in Diversity Outbred mice reveals Hydin as a novel pain gene. Mamm Genome. 2014;25(5-6):211-22.

168. Grant KA, Leng X, Green HL, Szeliga KT, Rogers LS, Gonzales SW. Drinking typography established by scheduled induction predicts chronic heavy drinking in a monkey model of ethanol self-administration. Alcohol Clin Exp Res. 2008;32(10):1824-38.
169. Helms CM, Park B, Grant KA. Adrenal steroid hormones and ethanol self-administration in male rhesus macaques. Psychopharmacology (Berl). 2014;231(17):3425-36.
170. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic acids research. 2019;47(D1):D1005-d12.
171. Munger SC, Raghupathy N, Choi K, Simons AK, Gatti DM, Hinerfeld DA, et al. RNA-Seq Alignment to Individualized Genomes Improves Transcript Abundance Estimates in Multiparent Populations. Genetics. 2014;198(1):59-73.
172. 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;108(17):7119-24.
173. White JJ, Arancillo M, King A, Lin T, Miterko LN, Gebre SA, et al. Pathogenesis of severe ataxia and tremor without the typical signs of neurodegeneration. Neurobiol Dis. 2016;86:86-98.
174. Shimobayashi E, Kapfhammer JP. Calcium Signaling, PKC Gamma, IP3R1 and CAR8 Link Spinocerebellar Ataxias and Purkinje Cell Dendritic Development. Curr Neuropharmacol. 2018;16(2):151-9.
175. Shimobayashi E, Wagner W, Kapfhammer JP. Carbonic Anhydrase 8 Expression in Purkinje Cells Is Controlled by PKCgamma Activity and Regulates Purkinje Cell Dendritic Growth. Mol Neurobiol. 2016;53(8):5149-60.
176. Artegiani B, de Jesus Domingues AM, Bragado Alonso S, Brandl E, Massalini S, Dahl A, et al. Tox: a multifunctional transcription factor and novel regulator of mammalian corticogenesis. Embo j. 2015;34(7):896-910.
177. Frederick JP, Tafari AT, Wu SM, Megosh LC, Chiou ST, Irving RP, et al. A role for a lithium-inhibited Golgi nucleotidase in skeletal development and sulfation. Proc Natl Acad Sci U S A. 2008;105(33):11605-12.
178. Tarantino LM, McClearn GE, Rodriguez LA, Plomin R. Confirmation of quantitative trait loci for alcohol preference in mice. Alcohol Clin Exp Res. 1998;22(5):1099-105.
179. Henderson JR, Macalma T, Brown D, Richardson JA, Olson EN, Beckerle MC. The LIM protein, CRP1, is a smooth muscle marker. Dev Dyn. 1999;214(3):229-38.
180. Boczkowska M, Yurtsever Z, Rebowski G, Eck MJ, Dominguez R. Crystal Structure of Leiomodin 2 in Complex with Actin: A Structural and Functional Reexamination. Biophys J. 2017;113(4):889-99.
181. Martinez-Lopez MJ, Alcantara S, Mascaro C, Perez-Branguli F, Ruiz-Lozano P, Maes T, et al. Mouse neuron navigator 1, a novel microtubule-associated protein involved in neuronal migration. Mol Cell Neurosci. 2005;28(4):599-612.
182. Lankford KL, Letourneau PC. Roles of actin filaments and three second-messenger systems in short-term regulation of chick dorsal root ganglion neurite outgrowth. Cell Motil Cytoskeleton. 1991;20(1):7-29.

183. Cruz-Rivera YE, Perez-Morales J, Santiago YM, Gonzalez VM, Morales L, Cabrera-Rios M, et al. A Selection of Important Genes and Their Correlated Behavior in Alzheimer's Disease. J Alzheimers Dis. 2018;65(1):193-205.
184. Bryzgalov LO, Korbolina EE, Brusentsov, II, Leberfarb EY, Bondar NP, Merkulova TI. Novel functional variants at the GWAS-implicated loci might confer risk to major depressive disorder, bipolar affective disorder and schizophrenia. BMC Neurosci. 2018;19(Suppl 1):22.
185. Savitz J, Frank MB, Victor T, Bebak M, Marino JH, Bellgowan PS, et al. Inflammation and neurological disease-related genes are differentially expressed in depressed patients with mood disorders and correlate with morphometric and functional imaging abnormalities. Brain Behav Immun. 2013;31:161-71.
186. Huang P, Cai Y, Zhao B, Cui L. Roles of NUCKS1 in Diseases: Susceptibility, Potential Biomarker, and Regulatory Mechanisms. Biomed Res Int. 2018;2018:7969068.
187. Meyer SU, Krebs S, Thirion C, Blum H, Krause S, Pfaffl MW. Tumor Necrosis Factor Alpha and Insulin-Like Growth Factor 1 Induced Modifications of the Gene Expression Kinetics of Differentiating Skeletal Muscle Cells. PLoS One. 2015;10(10):e0139520.
188. Moloney GM, van Oeffelen W, Ryan FJ, van de Wouw M, Cowan C, Claesson MJ, et al. Differential gene expression in the mesocorticolimbic system of innately high- and low-impulsive rats. Behav Brain Res. 2019;364:193-204.
189. de Monasterio-Schrader P, Patzig J, Mobius W, Barrette B, Wagner TL, Kusch K, et al. Uncoupling of neuroinflammation from axonal degeneration in mice lacking the myelin protein tetraspanin-2. Glia. 2013;61(11):1832-47.
190. Morinville A, Martin S, Lavallee M, Vincent JP, Beaudet A, Mazella J. Internalization and trafficking of neurotensin via NTS3 receptors in HT29 cells. Int J Biochem Cell Biol. 2004;36(11):2153-68.
191. Skeldal S, Sykes AM, Glerup S, Matusica D, Palstra N, Autio H, et al. Mapping of the interaction site between sortilin and the p75 neurotrophin receptor reveals a regulatory role for the sortilin intracellular domain in p75 neurotrophin receptor shedding and apoptosis. J Biol Chem. 2012;287(52):43798-809.
192. Nielsen MS, Gustafsen C, Madsen P, Nyengaard JR, Hermey G, Bakke O, et al. Sorting by the cytoplasmic domain of the amyloid precursor protein binding receptor SorLA. Mol Cell Biol. 2007;27(19):6842-51.
193. Ruan CS, Yang CR, Li JY, Luo HY, Bobrovskaya L, Zhou XF. Mice with Sort1 deficiency display normal cognition but elevated anxiety-like behavior. Exp Neurol. 2016;281:99-108.
194. Buttenschon HN, Elfving B, Nielsen M, Skeldal S, Kaas M, Mors O, et al. Exploring the sortilin related receptor, SorLA, in depression. J Affect Disord. 2018;232:260-7.
195. Capsoni S, Amato G, Vignone D, Criscuolo C, Nykjaer A, Cattaneo A. Dissecting the role of sortilin receptor signaling in neurodegeneration induced by NGF deprivation. Biochem Biophys Res Commun. 2013;431(3):579-85.
196. Reuter E, Weber J, Paterka M, Ploen R, Breiderhoff T, van Horssen J, et al. Role of Sortilin in Models of Autoimmune Neuroinflammation. J Immunol. 2015;195(12):5762-9.

197. Sanchez E, Bergareche A, Krebs CE, Gorostidi A, Makarov V, Ruiz-Martinez J, et al. SORT1 Mutation Resulting in Sortilin Deficiency and p75(NTR) Upregulation in a Family With Essential Tremor. ASN Neuro. 2015;7(4).
198. Belbin O, Morgan K, Medway C, Warden D, Cortina-Borja M, van Duijn CM, et al. The Epistasis Project: A Multi-Cohort Study of the Effects of BDNF, DBH, and SORT1 Epistasis on Alzheimer's Disease Risk. J Alzheimers Dis. 2019;68(4):1535-47.
199. Philtjens S, Van Mossevelde S, van der Zee J, Wauters E, Dillen L, Vandenbulcke M, et al. Rare nonsynonymous variants in SORT1 are associated with increased risk for frontotemporal dementia. Neurobiol Aging. 2018;66:181.e3-.e10.
200. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature genetics. 2012;44(12):1365-9.
201. Brečević L, Rinčić M, Krsnik Ž, Sedmak G, Hamid AB, Kosyakova N, et al. Association of new deletion/duplication region at chromosome 1p21 with intellectual disability, severe speech deficit and autism spectrum disorder-like behavior: an all-in approach to solving the DPYD enigma. Translational neuroscience. 2015;6(1):59-86.
202. Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, Black DL. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol Cell Biol. 2000;20(20):7463-79.
203. Doan RN, Bae BI, Cubelos B, Chang C, Hossain AA, Al-Saad S, et al. Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. Cell. 2016;167(2):341-54.e12.
204. Zhang M, Ergin V, Lin L, Stork C, Chen L, Zheng S. Axonogenesis Is Coordinated by Neuron-Specific Alternative Splicing Programming and Splicing Regulator PTBP2. Neuron. 2019;101(4):690-706.e10.
205. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011;477(7364):289-94.
206. Gill KJ, Boyle AE. Genetic basis for the psychostimulant effects of nicotine: a quantitative trait locus analysis in AcB/BcA recombinant congenic mice. Genes Brain Behav. 2005;4(7):401-11.
207. Gill K, Desaulniers N, Desjardins P, Lake K. Alcohol preference in AXB/BXA recombinant inbred mice: gender differences and gender-specific quantitative trait loci. Mamm Genome. 1998;9(12):929-35.
208. Doeppner TR, Herz J, Bahr M, Tonchev AB, Stoykova A. Zbtb20 Regulates Developmental Neurogenesis in the Olfactory Bulb and Gliogenesis After Adult Brain Injury. Mol Neurobiol. 2019;56(1):567-82.
209. Tonchev AB, Tuoc TC, Rosenthal EH, Studer M, Stoykova A. Zbtb20 modulates the sequential generation of neuronal layers in developing cortex. Mol Brain. 2016;9(1):65.
210. Nagao M, Ogata T, Sawada Y, Gotoh Y. Zbtb20 promotes astrocytogenesis during neocortical development. Nat Commun. 2016;7:11102.
211. Dong Q, Chen XY, Li GM. Effect of transcription factor ZBTB20 on mouse pituitary development. Genet Mol Res. 2015;14(4):17622-9.

212. Ren A, Zhang H, Xie Z, Ma X, Ji W, He DZ, et al. Regulation of hippocampus-dependent memory by the zinc finger protein Zbtb20 in mature CA1 neurons. J Physiol. 2012;590(19):4917-32.
213. Mitchelmore C, Kjaerulff KM, Pedersen HC, Nielsen JV, Rasmussen TE, Fisker MF, et al. Characterization of two novel nuclear BTB/POZ domain zinc finger isoforms. Association with differentiation of hippocampal neurons, cerebellar granule cells, and macroglia. J Biol Chem. 2002;277(9):7598-609.
214. Oscar-Berman M, Bowirrat A. Genetic influences in emotional dysfunction and alcoholism-related brain damage. Neuropsychiatr Dis Treat. 2005;1(3):211-29.
215. Roy B, Wang Q, Dwivedi Y. Long Noncoding RNA-Associated Transcriptomic Changes in Resiliency or Susceptibility to Depression and Response to Antidepressant Treatment. Int J Neuropsychopharmacol. 2018;21(5):461-72.
216. Davies MN, Krause L, Bell JT, Gao F, Ward KJ, Wu H, et al. Hypermethylation in the ZBTB20 gene is associated with major depressive disorder. Genome Biol. 2014;15(4):R56.
217. Ikeda M, Tomita Y, Mouri A, Koga M, Okochi T, Yoshimura R, et al. Identification of novel candidate genes for treatment response to risperidone and susceptibility for schizophrenia: integrated analysis among pharmacogenomics, mouse expression, and genetic case-control association approaches. Biol Psychiatry. 2010;67(3):263-9.
218. Ho KWD, Han S, Nielsen JV, Jancic D, Hing B, Fiedorowicz J, et al. Genome-wide association study of seasonal affective disorder. Transl Psychiatry. 2018;8(1):190.
219. Keller F, Levitt P. Developmental and regeneration-associated regulation of the limbic system associated membrane protein in explant cultures of the rat brain. Neuroscience. 1989;28(2):455-74.
220. Keller F, Rimvall K, Barbe MF, Levitt P. A membrane glycoprotein associated with the limbic system mediates the formation of the septo-hippocampal pathway in vitro. Neuron. 1989;3(5):551-61.
221. Pimenta AF, Levitt P. Characterization of the genomic structure of the mouse limbic system-associated membrane protein (Lsamp) gene. Genomics. 2004;83(5):790-801.
222. Heinla I, Leidmaa E, Kongi K, Pennert A, Innos J, Nurk K, et al. Gene expression patterns and environmental enrichment-induced effects in the hippocampi of mice suggest importance of Lsamp in plasticity. Front Neurosci. 2015;9:205.
223. Singh K, Lillevali K, Gilbert SF, Bregin A, Narvik J, Jayaram M, et al. The combined impact of IgLON family proteins Lsamp and Neurotrimin on developing neurons and behavioral profiles in mouse. Brain Res Bull. 2018;140:5-18.
224. Innos J, Koido K, Philips MA, Vasar E. Limbic system associated membrane protein as a potential target for neuropsychiatric disorders. Front Pharmacol. 2013;4:32.
225. Innos J, Philips MA, Leidmaa E, Heinla I, Raud S, Reemann P, et al. Lower anxiety and a decrease in agonistic behaviour in Lsamp-deficient mice. Behav Brain Res. 2011;217(1):21-31.
226. Innos J, Leidmaa E, Philips MA, Sutt S, Alttoa A, Harro J, et al. Lsamp(-)/(-) mice display lower sensitivity to amphetamine and have elevated 5-HT turnover. Biochem Biophys Res Commun. 2013;430(1):413-8.

227. Ji Z, Zhang G, Chen L, Li J, Yang Y, Cha C, et al. Spastin Interacts with CRMP5 to Promote Neurite Outgrowth by Controlling the Microtubule Dynamics. Dev Neurobiol. 2018;78(12):1191-205.
228. Zhang C, Li D, Ma Y, Yan J, Yang B, Li P, et al. Role of spastin and protrudin in neurite outgrowth. J Cell Biochem. 2012;113(7):2296-307.
229. Solowska JM, D'Rozario M, Jean DC, Davidson MW, Marenda DR, Baas PW. Pathogenic mutation of spastin has gain-of-function effects on microtubule dynamics. J Neurosci. 2014;34(5):1856-67.
230. Zempel H, Mandelkow EM. Tau missorting and spastin-induced microtubule disruption in neurodegeneration: Alzheimer Disease and Hereditary Spastic Paraplegia. Mol Neurodegener. 2015;10:68.
231. Fernandez JR, Vogler GP, Tarantino LM, Vignetti S, Plomin R, McClearn GE. Sex-exclusive quantitative trait loci influences in alcohol-related phenotypes. Am J Med Genet. 1999;88(6):647-52.
232. Boyle AE, Gill K. Sensitivity of AXB/BXA recombinant inbred lines of mice to the locomotor activating effects of cocaine: a quantitative trait loci analysis. Pharmacogenetics. 2001;11(3):255-64.
233. Fritschy JM, Harvey RJ, Schwarz G. Gephyrin: where do we stand, where do we go? Trends Neurosci. 2008;31(5):257-64.
234. Egger G, Roetzer KM, Noor A, Lionel AC, Mahmood H, Schwarzbraun T, et al. Identification of risk genes for autism spectrum disorder through copy number variation analysis in Austrian families. Neurogenetics. 2014;15(2):117-27.
235. Lionel AC, Vaags AK, Sato D, Gazzellone MJ, Mitchell EB, Chen HY, et al. Rare exonic deletions implicate the synaptic organizer Gephyrin (GPHN) in risk for autism, schizophrenia and seizures. Hum Mol Genet. 2013;22(10):2055-66.
236. Dejanovic B, Lal D, Catarino CB, Arjune S, Belaidi AA, Trucks H, et al. Exonic microdeletions of the gephyrin gene impair GABAergic synaptic inhibition in patients with idiopathic generalized epilepsy. Neurobiol Dis. 2014;67:88-96.
237. Enoch MA, Zhou Z, Kimura M, Mash DC, Yuan Q, Goldman D. GABAergic gene expression in postmortem hippocampus from alcoholics and cocaine addicts; corresponding findings in alcohol-naive P and NP rats. PLoS One. 2012;7(1):e29369.
238. Carayol J, Sacco R, Tores F, Rousseau F, Lewin P, Hager J, et al. Converging evidence for an association of ATP2B2 allelic variants with autism in male subjects. Biol Psychiatry. 2011;70(9):880-7.
239. Prandini P, Pasquali A, Malerba G, Marostica A, Zusi C, Xumerle L, et al. The association of rs4307059 and rs35678 markers with autism spectrum disorders is replicated in Italian families. Psychiatr Genet. 2012;22(4):177-81.
240. Schultz JM, Yang Y, Caride AJ, Filoteo AG, Penheiter AR, Lagziel A, et al. Modification of human hearing loss by plasma-membrane calcium pump PMCA2. N Engl J Med. 2005;352(15):1557-64.
241. Pottorf WJ, 2nd, Johanns TM, Derrington SM, Strehler EE, Enyedi A, Thayer SA. Glutamate-induced protease-mediated loss of plasma membrane Ca2+ pump activity in rat hippocampal neurons. J Neurochem. 2006;98(5):1646-56.
242. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE. Mouse Genome Database (MGD) 2019. Nucleic acids research. 2019;47(D1):D801-d6.

243. Griswold AJ, Ma D, Cukier HN, Nations LD, Schmidt MA, Chung RH, et al. Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways. Hum Mol Genet. 2012;21(15):3513-23.
244. Volk D, Austin M, Pierri J, Sampson A, Lewis D. GABA transporter-1 mRNA in the prefrontal cortex in schizophrenia: decreased expression in a subset of neurons. Am J Psychiatry. 2001;158(2):256-65.
245. Pesz K, Pienkowski VM, Pollak A, Gasperowicz P, Sykulski M, Kosinska J, et al. Phenotypic consequences of gene disruption by a balanced de novo translocation involving SLC6A1 and NAA15. Eur J Med Genet. 2018;61(10):596-601.
246. Mattison KA, Butler KM, Inglis GAS, Dayan O, Boussidan H, Bhambhani V, et al. SLC6A1 variants identified in epilepsy patients reduce gamma-aminobutyric acid transport. Epilepsia. 2018;59(9):e135-e41.
247. Enoch MA, Hodgkinson CA, Shen PH, Gorodetsky E, Marietta CA, Roy A, et al. GABBR1 and SLC6A1, Two Genes Involved in Modulation of GABA Synaptic Transmission, Influence Risk for Alcoholism: Results from Three Ethnically Diverse Populations. Alcohol Clin Exp Res. 2016;40(1):93-101.
248. Hu JH, Ma YH, Yang N, Mei ZT, Zhang MH, Fei J, et al. Up-regulation of gamma-aminobutyric acid transporter I mediates ethanol sensitivity in mice. Neuroscience. 2004;123(4):807-12.
249. Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. The Lancet Psychiatry. 2016;3(8):760-73.
250. Boyle AE, Gill KJ. Genetic analysis of the psychostimulant effects of nicotine in chromosome substitution strains and F2 crosses derived from A/J and C57BL/6J progenitors. Mamm Genome. 2009;20(1):34-42.
251. Hood HM, Metten P, Crabbe JC, Buck KJ. Fine mapping of a sedative-hypnotic drug withdrawal locus on mouse chromosome 11. Genes Brain Behav. 2006;5(1):1-10.
252. DeWire SM, Ahn S, Lefkowitz RJ, Shenoy SK. Beta-arrestins and cell signaling. Annu Rev Physiol. 2007;69:483-510.
253. Lefkowitz RJ, Rajagopal K, Whalen EJ. New roles for beta-arrestins in cell signaling: not just for seven-transmembrane receptors. Mol Cell. 2006;24(5):643-52.
254. Li H, Tao Y, Ma L, Liu X, Ma L. beta-Arrestin-2 inhibits preference for alcohol in mice and suppresses Akt signaling in the dorsal striatum. Neurosci Bull. 2013;29(5):531-40.
255. Jacob JC, Sakakibara K, Mischel RA, Henderson G, Dewey WL, Akbarali HI. Ethanol Reversal of Oxycodone Tolerance in Dorsal Root Ganglia Neurons. Mol Pharmacol. 2018;93(5):417-26.
256. O'Brien WT, Huang J, Buccafusca R, Garskof J, Valvezan AJ, Berry GT, et al. Glycogen synthase kinase-3 is essential for beta-arrestin-2 complex formation and lithium-sensitive behaviors in mice. J Clin Invest. 2011;121(9):3756-62.
257. Wang Y, Jin L, Song Y, Zhang M, Shan D, Liu Y, et al. beta-arrestin 2 mediates cardiac ischemia-reperfusion injury via inhibiting GPCR-independent cell survival signalling. Cardiovasc Res. 2017;113(13):1615-26.
258. Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, et al. Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. Am J Hum Genet. 2016;98(1):58-74.

259. Tang BL. ADAMTS: a novel family of extracellular matrix proteases. Int J Biochem Cell Biol. 2001;33(1):33-44.
260. Steiner JL, Pruznak AM, Navaratnarajah M, Lang CH. Alcohol Differentially Alters Extracellular Matrix and Adhesion Molecule Expression in Skeletal Muscle and Heart. Alcohol Clin Exp Res. 2015;39(8):1330-40.
261. Crespo-Facorro B, Prieto C, Sainz J. Schizophrenia gene expression profile reverted to normal levels by antipsychotics. Int J Neuropsychopharmacol. 2014;18(4).
262. Ferrer-Ferrer M, Dityatev A. Shaping Synapses by the Neural Extracellular Matrix. Front Neuroanat. 2018;12:40.
263. The Genomes Project C. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
264. Tyler RE, Kim SW, Guo M, Jang YJ, Damadzic R, Stodden T, et al. Detecting neuroinflammation in the brain following chronic alcohol exposure in rats: A comparison between in vivo and in vitro TSPO radioligand binding. Eur J Neurosci. 2019.
265. Kohno M, Link J, Dennis LE, McCready H, Huckans M, Hoffman WF, et al. Neuroinflammation in addiction: A review of neuroimaging studies and potential immunotherapies. Pharmacol Biochem Behav. 2019;179:34-42.
266. Warton K, Lin V, Navin T, Armstrong NJ, Kaplan W, Ying K, et al. Methylation-capture and Next-Generation Sequencing of free circulating DNA from human plasma. BMC Genomics. 2014;15(1):476.
267. George O, Koob GF. Control of craving by the prefrontal cortex. Proc Natl Acad Sci U S A. 2013;110(11):4165-6.
268. Choi K. EMASE: Estimation-Maximization algorithm for Allele Specific Expression. GitHub Inc. ; 2016.
269. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30(15):2114-20.
270. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013;29(1):15-21.
271. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics (Oxford, England). 2015;31(2):166-9.
272. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015;43(7):e47.
273. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
274. Mignogna KM, Bacanu SA, Riley BP, Wolen AR, Miles MF. Cross-species alcohol dependence-associated gene networks: Co-analysis of mouse brain gene expression and human genome-wide association data. bioRxiv. 2018.
275. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics. 2016;48(10):1279-83.
276. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nature genetics. 2016;48(10):1284-7.

277. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics (Oxford, England). 2016;32(9):1423-6.
278. Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. PLoS Comput Biol. 2016;12(1):e1004714.
279. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic acids research. 2009;37(Web Server issue):W305-11.
280. Liu J, Dietz K, DeLoyht JM, Pedre X, Kelkar D, Kaur J, et al. Impaired adult myelination in the prefrontal cortex of socially isolated mice. Nat Neurosci. 2012;15(12):1621-3.
281. Alexiou A, Nizami B, Khan FI, Soursou G, Vairaktarakis C, Chatzichronis S, et al. Mitochondrial Dynamics and Proteins Related to Neurodegenerative Diseases. Curr Protein Pept Sci. 2018;19(9):850-7.
282. Revett TJ, Baker GB, Jhamandas J, Kar S. Glutamate system, amyloid ß peptides and tau protein: functional interrelationships and relevance to Alzheimer disease pathology. J Psychiatry Neurosci. 2013;38(1):6-23.
283. Nennig SE, Schank JR. The Role of NFkB in Drug Addiction: Beyond Inflammation. Alcohol Alcohol. 2017;52(2):172-9.
284. Zhan X, Stamova B, Sharp FR. Lipopolysaccharide Associates with Amyloid Plaques, Neurons and Oligodendrocytes in Alzheimer's Disease Brain: A Review. Front Aging Neurosci. 2018;10:42.
285. Orso R, Creutzberg KC, Centeno-Silva A, Carapecos MS, Levandowski ML, Wearick-Silva LE, et al. NFkappaB1 and NFkappaB2 gene expression in the prefrontal cortex and hippocampus of early life stressed mice exposed to cocaine-induced conditioned place preference during adolescence. Neurosci Lett. 2017;658:27-31.
286. Thangaraj A, Periyasamy P, Guo ML, Chivero ET, Callen S, Buch S. Mitigation of cocaine-mediated mitochondrial damage, defective mitophagy and microglial activation by superoxide dismutase mimetics. Autophagy. 2019:1-24.
287. Venkataraman A, Kalk N, Sewell G, C WR, Lingford-Hughes A. Alcohol and Alzheimer's Disease-Does Alcohol Dependence Contribute to Beta-Amyloid Deposition, Neuroinflammation and Neurodegeneration in Alzheimer's Disease? Alcohol Alcohol. 2017;52(2):158.
288. Hota SK, Bruneau BG. ATP-dependent chromatin remodeling during mammalian development. Development. 2016;143(16):2882-97.
289. Monden T, Kishi M, Hosoya T, Satoh T, Wondisford FE, Hollenberg AN, et al. p120 acts as a specific coactivator for 9-cis-retinoic acid receptor (RXR) on peroxisome proliferator-activated receptor-gamma/RXR heterodimers. Mol Endocrinol. 1999;13(10):1695-703.
290. Andersen LB, Ballester R, Marchuk DA, Chang E, Gutmann DH, Saulino AM, et al. A conserved alternative splice in the von Recklinghausen neurofibromatosis (NF1) gene produces two neurofibromin isoforms, both of which have GTPase-activating protein activity. Mol Cell Biol. 1993;13(1):487-95.
291. Su X, Kong C, Stahl PD. GAPex-5 mediates ubiquitination, trafficking, and degradation of epidermal growth factor receptor. J Biol Chem. 2007;282(29):21278-84.

292. Menon AG, Gusella JF, Seizinger BR. Progress toward the isolation and characterization of the genes causing neurofibromatosis. Brain Pathol. 1990;1(1):33-40.
293. Zheng X, Dumitru R, Lackford BL, Freudenberg JM, Singh AP, Archer TK, et al. Cnot1, Cnot2, and Cnot3 maintain mouse and human ESC identity and inhibit extraembryonic differentiation. Stem Cells. 2012;30(5):910-22.
294. Pinheiro I, Margueron R, Shukeir N, Eisold M, Fritzsch C, Richter FM, et al. Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. Cell. 2012;150(5):948-60.
295. He N, Liu M, Hsu J, Xue Y, Chou S, Burlingame A, et al. HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. Mol Cell. 2010;38(3):428-38.
296. Park Y, Yoon SK, Yoon JB. TRIP12 functions as an E3 ubiquitin ligase of APP-BP1. Biochem Biophys Res Commun. 2008;374(2):294-8.
297. Lee IC, Leung T, Tan I. Adaptor protein LRAP25 mediates myotonic dystrophy kinase-related Cdc42-binding kinase (MRCK) regulation of LIMK1 protein in lamellipodial F-actin dynamics. J Biol Chem. 2014;289(39):26989-7003.
298. Li L, Wan J, Sase S, Groger M, Pollak A, Korz V, et al. Protein kinases paralleling late-phase LTP formation in dorsal hippocampus in the rat. Neurochem Int. 2014;76:50-8.
299. Lazzaretti D, Tournier I, Izaurralde E. The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins. Rna. 2009;15(6):1059-66.
300. Jackman SL, Regehr WG. The Mechanisms and Functions of Synaptic Facilitation. Neuron. 2017;94(3):447-64.
301. Zhuang B, Su YS, Sockanathan S. FARP1 promotes the dendritic growth of spinal motor neuron subtypes through transmembrane Semaphorin6A and PlexinA4 signaling. Neuron. 2009;61(3):359-72.
302. Luo N, Li G, Li Y, Fan X, Wang Y, Ye X, et al. SAMD4B, a novel SAM-containing protein, inhibits AP-1-, p53- and p21-mediated transcriptional activity. BMB Rep. 2010;43(5):355-61.
303. Oda K, Shiratsuchi T, Nishimori H, Inazawa J, Yoshikawa H, Taketani Y, et al. Identification of BAIAP2 (BAI-associated protein 2), a novel human homologue of hamster IRSp53, whose SH3 domain interacts with the cytoplasmic domain of BAI1. Cytogenet Cell Genet. 1999;84(1-2):75-82.
304. Disanza A, Mantoani S, Hertzog M, Gerboth S, Frittoli E, Steffen A, et al. Regulation of cell shape by Cdc42 is mediated by the synergic actin-bundling activity of the Eps8-IRSp53 complex. Nat Cell Biol. 2006;8(12):1337-47.
305. Choe Y, Pleasure SJ, Mira H. Control of Adult Neurogenesis by Short-Range Morphogenic-Signaling Molecules. Cold Spring Harb Perspect Biol. 2015;8(3):a018887.
306. Crews FT, Nixon K. Mechanisms of neurodegeneration and regeneration in alcoholism. Alcohol Alcohol. 2009;44(2):115-27.
307. Imai S, Kano M, Nonoyama K, Ebihara S. Behavioral characteristics of ubiquitin-specific peptidase 46-deficient mice. PLoS One. 2013;8(3):e58566.
308. Massaly N, Frances B, Mouledous L. Roles of the ubiquitin proteasome system in the effects of drugs of abuse. Front Mol Neurosci. 2014;7:99.

309. Melendez RI, McGinty JF, Kalivas PW, Becker HC. Brain region-specific gene expression changes after chronic intermittent ethanol exposure and early withdrawal in C57BL/6J mice. Addict Biol. 2012;17(2):351-64.
310. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014;515(7527):355-64.
311. Alexander-Kaufman K, James G, Sheedy D, Harper C, Matsumoto I. Differential protein expression in the prefrontal white matter of human alcoholics: a proteomics study. Mol Psychiatry. 2006;11(1):56-65.
312. Gutala R, Wang J, Kadapakkam S, Hwang Y, Ticku M, Li MD. Microarray analysis of ethanol-treated cortical neurons reveals disruption of genes related to the ubiquitin-proteasome pathway and protein synthesis. Alcohol Clin Exp Res. 2004;28(12):1779-88.
313. Mabb AM, Ehlers MD. Arc ubiquitination in synaptic plasticity. Semin Cell Dev Biol. 2018;77:10-6.
314. Citri A, Alroy I, Lavi S, Rubin C, Xu W, Grammatikakis N, et al. Drug-induced ubiquitylation and degradation of ErbB receptor tyrosine kinases: implications for cancer therapy. Embo j. 2002;21(10):2407-17.