Genetic and environmental influences on alcohol behaviors: insight from the mouse transcriptome

by

Todd Michael Darlington

B.A. DePauw University, 2003

A dissertation submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Integrative Physiology
2013

This thesis entitled: “Genetic and environmental influences on alcohol behaviors: insight from the mouse transcriptome”
Written by Todd M. Darlington
has been approved for the Department of Integrative Physiology

_____
Marissa A. Ehringer, Ph.D.

_____
Monika R. Fleshner, Ph.D.

_____
Matthew B. McQueen, Sc.D.

_____
Robert L. Spencer, Ph.D.

_____
Richard A. Radcliffe, Ph.D.

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Darlington, Todd Michael (Ph.D., Integrative Physiology)

Genetic and environmental influences on alcohol behavior: insight from the mouse transcriptome

Thesis directed by Professor Marissa Ehringer

An estimated 5% of Americans currently have an alcohol use disorder (AUD), either abuse alcohol or are dependent, causing an incredible health and economic burden, as well as increased strain on family and friends. AUDs are approximately 50% heritable, and the purpose of these studies was to investigate aspects of genetic influence (initial sensitivity to alcohol) as well as environmental influence (exercise) on alcohol behaviors.

The Inbred Long Sleep (ILS) and Inbred Short Sleep (ISS) mouse strains are a model of genetic sensitivity to ethanol. We observed many genes differentially expressed between the two strains, including several in chromosomal regions previously shown to influence initial sensitivity to ethanol. Furthermore, utilizing Weighted Gene Co-expression Network Analysis (WGCNA) we identified several modules of co-expressed genes corresponding to strain differences. Several candidate genes were identified as well as functional categories and signaling pathways, which may play a role in the phenotypic differences between the two strains.

It has become apparent that different rewarding stimuli activate common reward pathways, with the potential to influence each other, i.e. hedonic substitution. We demonstrate that voluntary access to a running wheel substantially reduces the consumption and preference of ethanol in mice. Furthermore, we observed differential

gene expression of several candidate genes involved in regulating the mesolimbic dopaminergic pathway, which we hypothesized to be the focal point of hedonic substitution. These data suggest an important role for this pathway, and especially for *Bdnf* and *Slc18a2* in regulating hedonic substitution.

In order to identify additional candidate genes and pathways underlying hedonic substitution in the striatum, we quantitatively sequenced the striatal transcriptome of mice consuming ethanol, exercising, and doing both or neither, and identified differentially expressed genes and WGCNA co-expression modules. Interestingly, several genes and functional groups differentially expressed in response to exercise were previously identified in our study of ILS and ISS mice. This suggests that one way exercise might influence ethanol behavior is by sensitizing mice to the acute effects, thereby decreasing consumption.

In conclusion, baseline genetic differences contribute to differential sensitivity to ethanol. In addition, the environmental influence of exercise induces a transcriptional response, possibly altering the response to ethanol, and resulting in hedonic substitution.

ACKNOWLEDGMENTS

The following work could not have been completed without the exceptional guidance and support of my mentor, Dr. Marissa Ehringer. I feel fortunate to have gained such a wealth of knowledge on science, on navigating academe, and on balancing life with work.

I wish to thank Dr. Richard Radcliffe for his guidance when confronted with the difficult task of learning new technology, and for his generosity with data. My work has been built on the foundation of what he has taught me.

I would also like to thank the members of the Genetics of Substance Abuse Laboratory: Drs. Nicole Hoft, Holly Stephens, Amber Flora, Xavi Gallego, and Helen Kamens for being mentors, friends, and encyclopedic sources of information when I had questions; and Jill Miyamoto and Ryan Cox for handling so much of the daily work, and saving the day many times over. I would like to thank the other graduate students I have worked with: Will Horton, Whitney Melroy, Dan Howrigan, Riley McCarthy, and others who have shared my misery at times. Thanks to everyone at the Institute for Behavioral Genetics, as it really is a collaborative effort.

Most importantly, I would like to thank my family and friends for their support and encouragement: Mom for her unconditional support, challenging me intellectually and living as an example of the correct metric of success; Erin for suffering through this alongside me and always providing an alternate hypothesis as well as good music; Amy and Kelly for their encouragement and for all eight nieces and nephews; and Ashley, who I rely on for so much, as a colleague, role model, training partner, and best friend.

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

1.1 Significance to the general public

The consumption of alcohol has been widespread throughout human history. Consequently, alcohol use disorders (AUDs) have become a leading cause of preventable disease and death. The World Health Organization estimates that at least 76.3 million people worldwide have an AUD, and alcohol use is implicated in 3.8% of all deaths worldwide [1, 2]. In addition, alcohol abusers are at increased risk for a number of diseases, including gastro esophageal cancer, cirrhosis of the liver, and epilepsy [3]. According to the National Institute on Alcohol Abuse and Alcoholism (NIAAA) approximately 40% of all traffic crash fatalities involve alcohol and in 2004 over $235 billion was used or lost on health care and decreased productivity related to alcohol use disorders [2, 4]. Although AUDs pose a significant burden on society, there are relatively few treatment options available. Overall relapse rates remain high (between 60-80%) and additional approaches are needed [5, 6]. Research on the etiology of AUDs has resulted in a shift from the perception of the disease as a lack of willpower or morals to

what is now accepted as a true psychological disease with multiple genetic and physical causes. To more fully understand this complex disease, further research is needed to identify the genetic and environmental variables conferring risk [6].

1.2 Evidence for genetic influences on ethanol behavior

Alcoholism has long been known as a familial disease [7, 8], and there are numerous studies citing the likelihood of multiple additional affected family members given that one member of the family has the disease [9]. Recently, it has been estimated that the heritability of AUDs is close to 50% [10], meaning the additive genetic contribution from all genetic loci can explain half of the variance in AUDs. To elucidate which genes contribute the most to the genetic variance of AUDs, population studies have been conducted [11, 12]. Several genes have been identified that account for some of the heritability [13]. These large population studies as well as longitudinal family studies led by the Collaboration on the Genetics of Alcoholism (COGA) have identified many genes, including several gamma-aminobutyric acid (GABA) receptor subunits [14-18], taste receptor subunits [19], neuropeptide-Y [20], and nuclear factor kappa B (NF-κB) [21]. However, like many psychiatric diseases, the etiology of AUDs is extremely complex and these genes only explain a small portion of the variance. The strongest findings come from polymorphisms in the groups of genes involved in the metabolism of ethanol (alcohol dehydrogenase and aldehyde dehydrogenase), affecting the ability of individuals to consume ethanol comfortably and mainly occurs in East Asian populations [11, 22-24].

Further evidence for the genetic influence on ethanol consumption comes from

laboratory animal studies. The use of inbred strains of mice, genetically identical within strain, allows for the control of environmental influences while studying the effect of different genetic backgrounds. Studies comparing different mouse inbred strains demonstrate that different strains exhibit different ethanol behaviors, including consumption [25-27] and ethanol-induced loss of righting reflex [28, 29]. Furthermore, several genetic tools exist (i.e. recombinant inbred lines and gene knockout mice) which enable researchers to implicate genomic regions and candidate genes as important for a specific phenotype. An example of recombinant inbred lines, crossing C57BL/6 mice (B6) with DBA/2 mice (D2), creates an F_1 generation of genetically identical mice, each diploid allele consisting of a B6 and a D2 allele. Crossing the F_1 hybrids with each other results in genetic recombination, and produces an F_2 generation in which each offspring's genome consists on average 50% each of B6 and D2, but with differing haplotypes. Inbreeding of the F_2 hybrids results in multiple recombinant inbred strains of mice, each genetically distinct. These BxD recombinant inbred lines typically display a spectrum of intermediate phenotypes relative to the parent strains. Combining these phenotypes with genotypic information, it is possible to identify regions of the genome that contribute the most to the observed parental phenotype. The regions are called quantitative trait loci (QTLs). BxD recombinant inbred lines have been used to determine a number of ethanol-related QTLs including, but not limited to, ethanol metabolism (chromosome 17) [30] and ethanol preference (chromosomes 2 and 9) [31-37]. The most comprehensive list of QTLs is maintained by the Portland Alcohol Research Center (www.ohsu.edu/parc/by_phen.shtml). The development of knockout and transgenic mice allows the study of the individual contribution of a gene on a

particular phenotype. Hundreds of genes have been knocked out or over-expressed, and their effects on ethanol behaviors studied. A complete review is beyond the scope of this introduction, but notable genes include *Slc18a2* [38]*, Drd2* [39-41]*, Pdyn* [42]*, Slc6a3* [43]*,* and *Prkcg* [44]*,* discussed in the following chapters.

These results demonstrate that in both humans and in mice, there is a strong genetic component to ethanol related behaviors.

1.3 Sensitivity to ethanol and an animal model of sensitivity

Using familial history of AUDs as a proxy for genetic risk, Schuckit (1980) demonstrated that male subjects at risk for AUDs were less sensitive to the subjective effects of consuming ethanol than their peers with family histories of AUDs [45]. Family and longitudinal studies confirmed that an initial low response predicted future risk of developing AUDs [46-49]. In animals, the loss of righting reflex (LORR) due to a single intraperitoneal injection of ethanol is a measure of the baseline sensitivity to ethanol [28]. The Long Sleep (LS) and Short Sleep (SS) mouse strains were developed as a model to the sensitivity to ethanol [50]. Following generations of selection for differences in LORR, the LS mice lose their righting reflex for approximately 16 times longer than the SS mice, independent of minor differences in ethanol metabolism. These two strains, and their inbred descendants, the Inbred Long Sleep (ILS) and Inbred Short Sleep (ISS), differ in a number of phenotypes besides LORR, including ethanol consumption. The less sensitive ISS mice consume more ethanol than the ILS mice [51], in agreement with the observations of Schuckit. Recombinant inbred lines (LSxSS and LxS) have been used to identify QTLs related to LORR on chromosomes 1, 2, 3, 8,

11, and 15 [52-56], and candidate genes located within those QTLs, including *Rassf2* and *Myo1d* [57], have been identified. These studies provide additional evidence for the genetic component of risk for AUDs, and suggest a role for initial sensitivity to ethanol in conferring that risk.

1.4 Exercise as an environmental influence on ethanol consumption, i.e. hedonic substitution

McMillan (1978) was the first to report the behavioral interaction of exercise and ethanol. Rats bred to consume high quantities of ethanol (P rats) were allowed to voluntarily consume ethanol over 10 days, and subsequently given access to a running wheel. The study was designed to determine baseline levels of activity, which would then be compared with activity during ethanol withdrawal. However, when introduced to the running wheel, the rats decreased their ethanol intake by approximately 50% [58, 59]. Werme et al. (2002) showed that male Lewis rats allowed to voluntarily exercise during ethanol withdrawal consume significantly more ethanol upon re-introduction [60]. The authors concluded that there were convergent neurobiological pathways mediating both behaviors, and in particular the mechanism behind the ethanol deprivation effect. Ozburn et al (2008) provided additional evidence for hedonic substitution of exercise for ethanol. In their protocol using female B6 mice, repeated removal and re-introduction of both wheel access and 10% ethanol showed that exercise modifies patterns of ethanol consumption. Specifically, they observed reduced ethanol preference the first time a wheel was introduced, although this effect of exercise eventually diminished [61]. Hammer et al (2010) showed that male Syrian hamsters reduced ethanol consumption,

but not preference, when given access to a running wheel. Furthermore, the effect was shown to be reciprocal, since introduction of ethanol to hamsters with previous access to running wheels decreased the distance voluntarily run [62]. Recent work by Ehringer et al (2009) supports the hypothesis that reward provided by wheel-running may substitute or overlap at least in part with reward provided by ethanol, by showing significantly reduced ethanol preference and consumption in exercising mice compared to sedentary mice, and more so than in mice housed with a locked wheel. Ehringer et al (2009) also found no metabolic differences between exercising and sedentary mice [63]. These studies support the hypothesis that hedonic substitution of ethanol by exercise could be a useful approach for treating ethanol abuse and dependence [64].

In addition to studies using animal models, exercise has been shown to be effective for reducing ethanol consumption in humans. There are several reported studies in human populations supporting the idea of hedonic substitution for treatment of ethanol abuse [65]. Three studies used exercise intervention as a tool to decrease ethanol intake. Murphy et al (1986) showed that in otherwise healthy but heavy drinking college students, supervised exercise 3 times per week for 8 weeks significantly reduced ethanol consumption during the course of the study [66]. Also in heavy drinking college students, Correia et al (2005) showed a similar decrease in consumption, even though over the 4 weeks of the intervention the exercise group was unsupervised, and only instructed to exercise more [67]. In the third study, Werch et al (2010) reported that high school students decreased the quantity and frequency of ethanol consumption after participating in a 2 month behavioral intervention, which included encouragement to exercise more [68]. A seemingly conflicting report by Gutgesell et al (1996) used a

mailed questionnaire to assess the exercise and drinking behaviors of a group of runners and non-runners. They showed that male runners consumed more ethanol than controls; importantly though, among the subjects who reported a history of ethanol abuse, runners reported drinking less than non-runners [69]. Similarly, intense exercise has been shown to lead to a significant decrease cravings in recovering alcoholics [70]. These results suggest that exercise may be an effective tool for reducing ethanol consumption in human populations.

1.5 Utilizing the neural transcriptome to study the genetic and environmental influences on ethanol behavior

Variability in ethanol related behaviors across different strains of mice—each with their own genetic background—exists at the initial exposure to ethanol, suggesting that baseline transcriptional differences could account for much of the phenotypic variance. Recent advances in technology have allowed for increasingly accurate and detailed glimpses of the transcriptome. In the ILS and ISS strains, studies using hybridization arrays identified numerous differentially expressed genes in both whole brain [71] and sub-regions including the cerebellum [72], ventral midbrain, and striatum [73].

Seven studies in mice or rats have examined the changes in brain region-specific gene expression after exposure to ethanol, via either acute intraperitoneal injections [74, 75], or after periods of voluntary consumption [76-80]. Ethanol is a “dirty” drug, meaning it has many molecular targets both in the brain and periphery, so unsurprisingly few candidate genes are identified across multiple studies. To circumvent this issue, combining treatment groups [78] or using liberal False Discovery Rates [79]

have been used to identify changes in gene expression, but perhaps more elegant was the use of network analysis in addition to gene expression, where Mulligan et al (2011) identified roles for cell-type specific responses (astrocytes) and blood circulation in the response to ethanol consumption [80].

One study to date has examined the effect of exercise on the rat hippocampal transcriptome [81], which found many exercise responsive genes, including *Bdnf, Vegfa,* as well as an inward rectifying potassium channel, among others. However, a number of targeted gene expression studies have implicated genes involved in regulation of the mesolimbic dopaminergic pathway [82-84].

The following studies examine the genetic and environmental influences on ethanol behavior. In Chapter 2, we examine the whole brain and striatal transcriptomes of the ILS and ISS mice using RNA-Sequencing to provide high resolution expression data [85]. RNA-Sequencing is not hindered by the high background noise or saturation as are hybridization microarrays, which allows for a much higher dynamic range of expression values [86, 87]. RNA-Sequencing also provides sequences for abundantly expressed genes, including polymorphisms. In addition, the accuracy and dynamic range translate into improved network characteristics compared with microarrays, as the expression data can be used to construct networks of co-expressed genes [88]. Combining differential expression testing, sequence polymorphism identification, and co-expression networks allowed us to identify candidate genes and gene networks which may contribute to differential sensitivity to ethanol. Chapters 3 and 4 examine the transcriptional response to behavior invoking the hedonic substitution of exercise for ethanol. Initially we utilize a candidate gene approach in multiple brain regions, then

employ a transcriptome wide analysis in the striatum using RNA-sequencing to identify additional candidate genes and networks.

Figure 1.1 Overview of the components of Chapters 2-4.

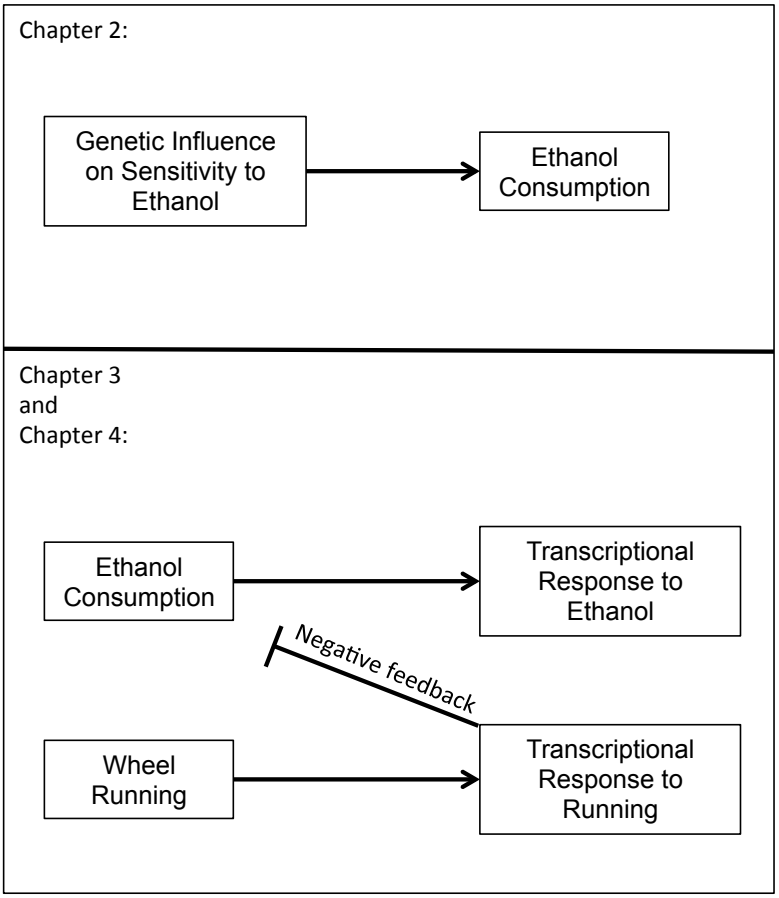

Figure 1.1 illustrates the topics covered in Chapters 2-4. Chapter 2 focuses on the genetic influence on sensitivity to ethanol, which according to the low level of response hypothesis [46] could influence observed differences in consumption [51]. Chapter 3 introduces the concept of hedonic substitution and examines the transcriptional response of candidate genes to wheel running and ethanol consumption, while Chapter 4 attempts to identify additional candidate genes using a transcriptome wide approach. Hedonic substitution is shown here as a transcriptional response that reduces ethanol consumption.

Chapter 2

Transcriptome analysis of Inbred Long Sleep and Inbred Short Sleep mice

Todd M Darlington, Marissa A Ehringer, Colin Larson,

Tzu L Phang, and Richard A Radcliffe

Genes, Brain, and Behavior. 2013, 12:263-274.

2.1 Abstract

Many studies have utilized the Inbred Long Sleep and Inbred Short Sleep mouse strains to model the genetic influence on initial sensitivity to ethanol. The mechanisms underlying this divergent phenotype are still not completely understood. In this study, we attempt to identify genes that are differentially expressed between these two strains and to identify baseline networks of co-expressed genes, which may provide insight regarding their phenotypic differences. We examined the whole brain and striatal transcriptomes of both strains, using next generation RNA sequencing techniques. Many genes were differentially expressed between strains, including several in chromosomal regions previously shown to influence initial sensitivity to ethanol. These results are in concordance with a similar sample of striatal transcriptomes measured using microarrays. In addition to the higher dynamic range, RNA-Seq is not hindered by high background noise or polymorphisms in probesets as with microarray technology, and we are able to analyze exome sequence of abundantly expressed genes. Furthermore, utilizing Weighted Gene Co-expression Network Analysis (WGCNA) we identified several modules of co-expressed genes corresponding to strain differences. Several candidate genes were identified, including protein phosphatase 1 regulatory unit 1b (*Ppp1r1b*), prodynorphin (*Pdyn*), proenkephalin (*Penk*), ras association (RalGDS/AF-6) domain family member 2 (*Rassf2*), myosin 1d (*Myo1d*), and transthyretin (*Ttr*). In addition, we propose a role for potassium channel activity as well as map kinase signaling in the observed phenotypic differences between the two strains.

2.2 Introduction

The heritability of alcohol use disorders, estimated to be approximately 0.5, suggests that genetics plays an important role in determining an individual's risk [10]. One possibility for how this risk manifests itself is in first response to alcohol [45], where it was demonstrated that a low level of response to alcohol is a strong predictor of future alcohol use disorders [46, 47]. In animals, measures of acute ethanol response from a single intra-peritoneal injection include: ethanol-stimulated activity, metabolism, hypothermia, ataxia, and loss of righting reflex (LORR). The Inbred Long Sleep (ILS) and Inbred Short Sleep (ISS) mouse strains were selected for differences in LORR and show a large phenotypic divergence [50]. Since this phenotype is present in ethanol-naïve animals, it is likely that genetically mediated differences in baseline gene expression could account for much of this phenotypic difference.

The ILS and ISS mice have been extensively studied, and are phenotypically different beyond ethanol-induced LORR [89-91], for example, the strains differ in ethanol preference with the ISS mice consuming more ethanol than the ILS mice [51]. The underlying genetics of these quantitative traits have been explored successfully using recombinant panels of mice to identify regions of interest likely involved in LORR (*Lore* QTLs) on chromosomes 1, 2, 3, 8, 11, and 15 [52-56, 92]. Genes in these regions were sequenced to find polymorphisms that may contribute to the observed phenotypes, and fifteen genes with coding sequence differences were identified [93]. Further, gene expression studies, in both whole brain [71], and cerebellum [72] identified many differentially expressed genes (DEGs) between the strains. Maclaren

identified several DEGs within *Lore* QTL regions with promoter region sequence differences [57].

The current study utilized Next Generation RNA Sequencing (RNA-Seq) technology to investigate baseline gene expression differences between these two strains. RNA-Seq produces millions of short reads which, when mapped back to the genome, provide a measure of gene expression as well as strain-specific sequence, at least for abundantly expressed genes. It provides a higher level of resolution of gene expression than is possible with hybridization microarrays. A high level of background noise, typical with microarrays, does not limit RNA-Seq [86, 87]. RNA-Seq has been shown to improve network characteristics compared to microarrays [88]. The purpose of this study is to identify both DEGs and networks of co-expressed genes for future study of initial response to alcohol and risk of alcohol use disorders. While priority will be given to genes previously identified in alcohol or drug studies, we will use multiple bioinformatics resources to filter candidate genes depending on differential expression, sequence differences, genome locations, and co-expression with other candidate genes.

2.3 Materials and Methods

2.3.1 Statement on animal care

This study was conducted with approval from the Institutional Animal Care and Use Committee at the University of Colorado Health Sciences Center (Denver, Colorado) following guidelines established by the Office of Laboratory Animal Welfare. All possible measures were taken to minimize animal discomfort.

2.3.2 RNA extraction

Mice were bred and housed at the specific pathogen free facility at the Institute for Behavioral Genetics (University of Colorado, Boulder) under a 12-hour light/dark cycle with *ad libitum* access to food and water. On post-natal day 60, twelve ethanol-naïve adult male mice (n=6/strain) were sacrificed by cervical dislocation and whole brains were removed. Six brains (n=3/strain) were further dissected to isolate the striatum. Total RNA, from whole brains (WB, n=6, 3/strain) and striatum (ST, n=6, 3/strain) was extracted using RNeasy midi kits (Qiagen, Valencia, California), and quantity and quality were determined using a NanoDrop™ spectrophotometer (Thermo Fisher Scientific, Wilmington, Delaware) and Agilent 2100 BioAnalyzer™ (Agilent Technologies, Santa Clara, California). Ratios of absorbance at 260nm and 280nm were shown to be excellent (>1.8). RNA Integrity scores were also shown to be excellent (>8.0).

2.3.3 Library preparation

The preparation of the cDNA library for RNA-Sequencing was conducted according to Illumina (San Diego, California) protocol for quantitative RNA Sequencing on the Genome Analyzer II (GAII) platform. Starting with 10 µg total RNA for each RNA sample, the samples were enriched for poly-A RNA using Sera Mag Magnetic Oligo(dT) Beads™. The poly-A enriched RNA samples were then fragmented with a 3M NaOAc solution at 94°C for 5 minutes. The samples were reverse transcribed with random primers, and end repair was performed with T4 and Klenow DNA polymerase. Double

stranded Illumina adaptors, with a single thymine overhang, were ligated to the ends of the cDNA fragments by first adding a single adenine to each 3' end of the cDNA. Next, 200bp fragments were selected by agarose gel electrophoresis and subsequent gel extraction with Qiagen Gel Purification kits. Libraries were enriched with 15 cycles of PCR, and purified using QIAquick PCR Purification kits (Qiagen). Each cDNA library was run on one GAII lane sequencing to 36bp.

2.3.4 Alignment

Raw 36 nucleotide reads were trimmed to 28nt due to inherent decrease in quality score toward the 3' end [94]. Reads were mapped to the mouse reference genome (mm9, Ensembl) using TopHat (v1.2.0, http://tophat.cbcb.umd.edu) [95]. TopHat first maps reads using Bowtie (v0.12.7, http://bowtie-bio.sourceforge.net/) [96] alignment software, which utilizes a Burrows-Wheeler index of the mouse genome (obtained from Bowtie source webpage, http://bowtie-bio.sourceforge.net/) to rapidly align short reads. TopHat then uses the resulting read pileup to deduce likely exon/intron boundaries, and identifies reads aligning across boundaries. Reads with up to 2 mismatches were allowed, and reads were removed if they aligned to more than 10 places in the genome. Visualization of read pileups was done using the Integrated Genomics Viewer (IGV v2.1, www.broadinstitute.org/igv) [97].

2.3.5 Transcript assembly, quantification, and differential expression testing

To assemble transcripts and estimate abundance, output from TopHat and the annotated reference genome (mm9, Ensembl) was analyzed using Cufflinks (v2.0.2,

http://cufflinks.cbcb.umd.edu/) [98] to construct the minimum number of transcripts that explain the maximum number of reads. Since the sequenced sample had been enriched for poly-A mRNA transcripts, a mask file was used to discriminate against alignments in rRNA, tRNA, and small RNA genes. Once transcripts were assembled, their abundances were estimated by counting the number of aligned reads contained in the transcript, and normalizing both to the size of transcript and to the total number of aligned reads in the sample (fragments per kilobase exon per million mapped fragments, FPKM). Cuffcompare was then used to compile the set of transcripts from each group, and each transcript was tested for differential expression using Cuffdiff. Data for the four groups of three samples (ILS/WB, ILS/ST, ISS/WB, and ISS/ST) were input into Cuffdiff to calculate each pairwise comparison of gene expression. Cuffdiff outputs estimates of the Jensen-Shannon divergence of each pair to determine statistical significance. Due to the exploratory nature of this study, we applied a less stringent correction for multiple testing, using a False Discovery Rate (FDR=0.1). Since the Cuffdiff minimum threshold of 1000 reads allows inclusion of intronic reads, reads aligning to close neighbors, and/or genes contained within an intron, we wanted to ensure that we only included reads which aligned within the exon structure, therefore we set a minimum expression level FPKM of at least 1 for genes to be included in subsequent analyses. Minimum thresholds have been employed in previous studies, and a minimum FPKM of 1 is consistent [99, 100]. In addition, using the Ensembl annotation information, we identified expressed genes (FPKM>1) with overlapping features, i.e. un-translated regions on opposite strands. Visual examination of each of these cases resulted in removal of 139 genes from further analysis.

2.3.6 Weighted Gene Co-expression Network Analysis (WGCNA)

Weighted gene co-expression networks were generated using the statistical program R (v2.11.1, www.r-project.org) and the WGCNA package (http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/) [101-103]. Cufflinks output from all twelve samples were used for a single WGCNA. Data were merged based on unique Ensembl Gene Id, and genes were excluded if no group reached an average FPKM≥1. Briefly, WGCNA first attempted to impute missing data using a k-nearest neighbors algorithm, then removed genes where imputation was impossible, and removed genes with no variance in expression values. Next, a signed similarity matrix was constructed with Equation 2.1.

$$S_{ij} = \frac{1 + cor(x_i, x_j)}{2} \quad (2.1)$$

This was converted to a weighted adjacency matrix by a power function (Equation 2.2), determined by a scale-free topology model (β=4).

$$a_{ij} = S_{ij}^{\beta} \quad (2.2)$$

Therefore, the adjacency matrix contained values from 0 to 1 for each gene, with 0, 0.5, and 1; signifying negative correlation (0-0.5), no correlation (0.5), and positive correlation (0.5-1). Adjacency was converted to topological overlap (Equation 2.3).

$$TOM_{ij} = \frac{\sum_u a_{iu} a_{uj} + a_{ij}}{\min(k_i, k_j) + 1 - a_{ij}} \quad (2.3)$$

Genes were clustered based on hierarchical clustering of topological overlap matrix-based dissimilarity, with the dynamic tree cutting algorithm cutreeDynamic, and the deepSplit option set to 4. Gene clusters with a minimum of 20 genes were identified

using a dynamic tree-cutting algorithm, which identified 21 gene clusters (modules). Similar gene modules were merged using the mergeCloseModules command, with a dissimilarity threshold of 0.1 (Pearson correlation greater than 0.9). Merging similar modules resulted in 16 remaining modules used in downstream analysis. Hub genes in each module were determined by ranking each gene by its module membership, calculated by WGCNA. Module robustness was tested in two ways. First, average module adjacencies were calculated and compared to the average adjacencies of randomly sampled “modules” of the same size. One thousand permutations of randomly sampled modules were generated. Modules were considered robust if average module adjacencies were significantly higher than the randomly generated modules. Second, the intramodular and extramodular connectivity of each module was calculated and scaled according to module size. Modules with higher scaled intramodular connectivity were considered robust.

2.3.7 Identification of relevant co-expression modules

To identify biologically relevant co-expression modules, we took the first principle component of each module, or module eigengene, using the moduleEigengenes command from the WGCNA R-package. Each module eigengene is representative of the gene expression levels for each module, if the module were reduced to a single gene. An analysis of variance of the resulting module eigengene values was used to identify module eigengenes different due to strain, region, or an interaction. Significant p-values were less than 0.05/16=0.003125. Each module was tested for enrichment of differentially expressed genes using a hypergeometric distribution function in R, and p-

values were corrected using the p.adjust function in R, utilizing the Benjamini-Hochberg method [104]. The set of differentially expressed genes had been determined using the Cufflinks package as described above, and genes were included if significant at FDR=0.1.

2.3.8 Bioinformatics analyses

The set of differentially expressed genes were tested for functional group over-representation with the Web-based gene set analysis toolkit (WebGestalt, http://bioinfo.vanderbilt.edu/webgestalt) [105, 106]. Functional groups based on Gene Ontology (GO) [107], Kyoto Encyclopedia of Genes and Genomes (KEGG) [108, 109], and WikiPathways [110, 111]. Over-represented *Lore* QTL regions were identified using a hypergeometric distribution function in R. *Cis*-regulation of differentially expressed and WGCNA module *Lore* QTL hub genes was determined using publicly available datasets at www.genenetwork.org. Expression QTLs were identified using two LxS datasets, hippocampus (Aug07) and prefrontal cortex (Aug06), as well as two BxD datasets, striatum (Dec10v2) and whole brain (Nov06). Peak LOD score for expression must occur within 10Mb of gene locus to have been considered *cis*-regulated. Furthermore, since multiple datasets were used to interrogate regulation of expression, and most datasets contained multiple probes for each gene, *cis*- peaks had to occur in the majority of all the probes and at least once in each dataset to be considered having evidence of *cis*-regulation. MicroRNA binding sites were identified from www.microrna.org, visualizing all miRNAs with good mirSVR scores. In addition, sets of differentially expressed genes and co-expression modules were tested for over-

representation of genes previously identified as being significantly differentially expressed (at least 3-fold higher) by cell type—neuron, astrocyte, or oligodendrocyte [112].

2.3.9 Identifying gene sequence differences

Cis-regulated differentially expressed genes in *Lore* QTL regions, as well as *Lore* QTL hub genes from WGCNA modules were visualized in IGV to identify sequence differences between strains. IGV incorporates annotated SNP information from dbSNP (build 128), which we used to classify SNPs as known or novel. In addition, genes sequenced previously [93] were visualized for confirmation of previous results.

2.3.10 Affymetrix microarray analysis

A reanalysis of previously published ILS/ISS striatal Affymetrix microarray results [73] was conducted as a validation study of the current RNA-seq DEG results. Briefly, striatal tissue was dissected and total RNA was isolated from 15 naïve mice from each strain. RNA was quantitatively pooled from 3 mice for a total of 5 microarray samples for each strain. RNA preparation, array hybridization (Affymetrix 430 v2.0), and array scanning was performed using standard procedures; details can be found in Radcliffe et al (2006).

Two probe masks were created and implemented to eliminate erroneous probes from calculations of transcript expression, thereby, increasing accuracy of expression estimates. Probe sequences were obtained directly from Affymetrix and aligned to the mouse genome (mm9) using BLAT [113]. First, individual probes that aligned to more

than one location or did not perfectly align were removed. Second, probes that targeted regions of the genome harboring SNPs were eliminated because an “expression” difference detected from these probes was more likely to represent differences in hybridization efficiency rather than true differences in RNA expression levels [114]. SNPs were identified from the current RNA-seq data using Partek Genomics Suite (v6.6; St. Louis, MO). We were less concerned about keeping probesets as ensuring that the retained probesets were of the highest quality possible. A liberal statistical criterion was thus used to test for significance of the SNPs (LOD>5.0) at the risk of increased type I errors for SNP identification, but at the same time, increased type II errors for probe removal, which we felt was acceptable in this case. Finally, probesets were required to consist of at least five probes. Following a global scaling procedure (average signal intensity of each array was set to a default target signal of 500), probe level normalization was performed using the Robust Multi-array Average method (RMA). Any RMA value that was less than 0.01 was converted to 0.01.

2.4 Results

2.4.1 Illumina GAII sequencing

Quantitative RNA Sequencing was completed on an Illumina GAII platform. Twelve samples total were sequenced, 6 each of whole brain (WB) and striatum (ST). Three samples from each region were from ILS mice, three from ISS mice. Whole brain data yielded short-read libraries of 12.7 and 13.1 million reads on average in ILS and ISS strains respectively. Striatum sequencing produced libraries of 26.9 and 26.5 million

reads on average in ILS and ISS strains (Table 2.1). Differences in library size are due to updates in Illumina software occurring between sequencing dates.

2.4.2 Alignment

Approximately 0.02% of low-complexity reads were discarded prior to alignment. Of the remaining reads, when alignment was constrained to 2 mismatches and 1 alignment, between 72 and 75% of reads aligned to the mouse genome. When constraints were relaxed to allow for up to 10 alignments, ~89% of reads were aligned. Over 70,000 (WB) and 80,000 (ST) unique exon-exon boundaries were identified (Table 2.1).

2.4.3 Differential expression

Using a minimum expression threshold of FPKM≥1 (in at least one sample) and a false discovery rate (FDR=0.1), 90 genes were differentially expressed between strains in the whole brain. In striatum, 336 genes were differentially expressed (Figure 2.1). Fifty-three genes were identified as differentially expressed in both data sets. Of those, 52 were differentially expressed in the same direction, while only one was higher in one strain compared to the other depending on region. Eight WB DEGs and 31 ST DEGs reside in previously identified *Lore* QTL regions. Noteworthy differences include 14 potassium channel subunit ST DEGs, previously identified candidate genes—ras

Table 2.1 Alignment statistics for each strain and region.

Region	Strain	# mice	Total reads[a]	Reads removed[b]	Unique hits[c]	Parameter hits[d]	# Exon junctions[e]
Striatum	ILS	3	26927097 ± 882830	4923 (0.0184%)	72.45%	88.89%	80214
	ISS	3	26466323 ± 1020682	5122 (0.0195%)	73.97%	89.63%	83274
Whole Brain	ILS	3	12786365 ± 1355373	3301 (0.0273%)	74.78%	89.17%	72643
	ISS	3	13130036 ± 481554	1853 (0.0141%)	74.37%	88.31%	71220

[a]Total number of short reads generated per group with standard deviation.
[b]Low complexity reads are filtered prior to any attempt to align.
[c]Percent of reads aligned to exactly one region of the genome.
[d]Percent of reads aligned when allowing for up to 10 alignments.
[e]Number of unique exon boundaries identified.

Figure 2.1 Differentially expressed genes in whole brain and striatum.

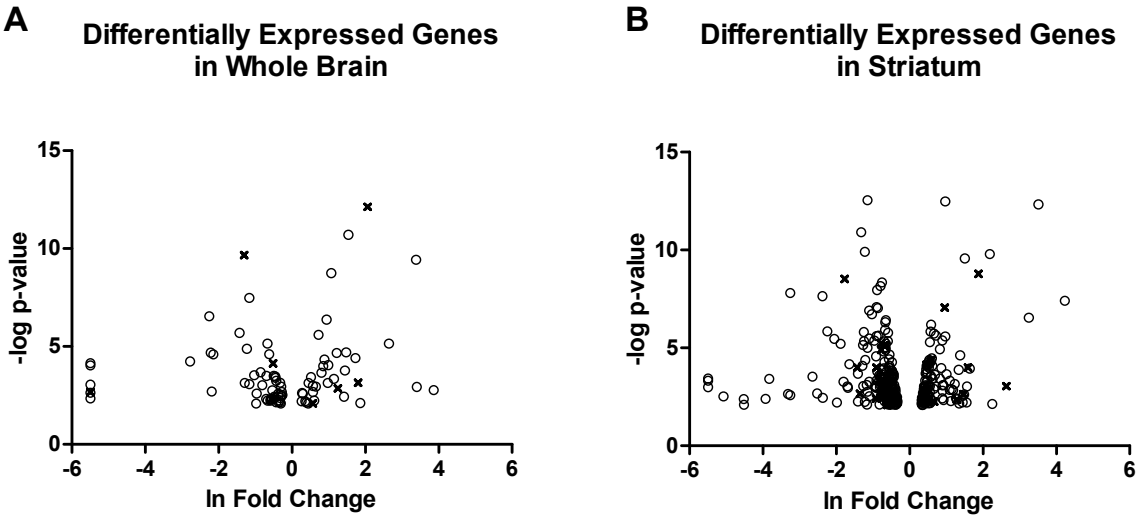

Figure 2.1 displays the distribution of differentially expressed genes between strain in Whole Brain (A) and Striatum (B) samples. The x-axis represents the natural log of the fold change, with positive values corresponding to higher expression in ILS mice, and negative values corresponding to higher expression in ISS mice. The y-axis represents the negative log of the p-value of the difference in expression, with more significant differences corresponding to higher numbers. Open circles (82 WB, 305 ST) represent genes significant at a False Discovery Rate (FDR) of 0.1. X's (8 WB, 31 ST) represent genes lying in Lore QTL regions.

association (RalGDS/AF-6) domain family member 2 (*Rassf2*) and myosin 1d (*Myo1d*), and genes previously implicated in ethanol/drug response phenotypes—protein phosphatase 1 regulatory unit 1b (*Ppp1r1b*), opioid peptide precursor genes prodynorphin (*Pdyn*) and proenkephalin (*Penk*), and transthyretin (*Ttr*).

Of the 336 DEG from the striatum, 297 had one or more valid probesets represented on the Affymetrix array. These Affymetrix probesets were tested for DE using one-way ANOVA (uncorrected; one-tail test). Over 90% of the Affymetrix probesets were expressed in the same direction as the RNA-seq DEG (Figure 2.2). Of these, 65.7% were DE at $p<0.05$, 10.8% were DE at a p value between 0.05 and 0.1, and the remainder were DE at $p>0.1$ (Figure 2.2).

2.4.4 Over-representation analysis of differentially expressed genes

Utilizing the online resource WebGestalt, GO and KEGG functional group, and chromosomal region over-representation was determined on the set of 90 differentially expressed genes in WB, and the 336 differentially expressed genes in ST, with the reference set of genes based on the total number of genes detected at FPKM≥1 and tested for differential expression (12,678 genes in WB, 12,395 in ST). The results are shown in Table 2.2. Briefly, the most significant functional groups represented in whole brain include groups related to ribosomes, extracellular regions, and the major histocompatibility protein complex (corrected group p-values range from 9.19×10^{-6} – 0.0285). In striatum, the most significant functional groups include those related to ribosomes, potassium channel activity, and signal transduction (corrected group p-values range from 3.45×10^{-6} – 0.0482).

Figure 2.2 Differentially expressed genes from striatum replicated in microarrays from Radcliffe et al (2006).

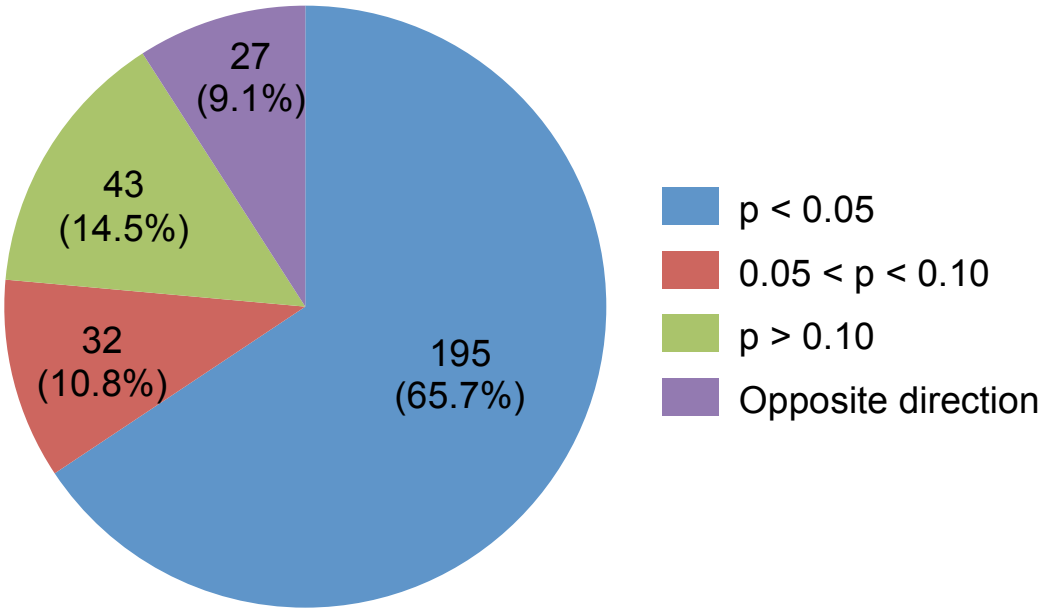

Figure 2.2 shows the number of differentially expressed genes from the striatum that are represented in microarray data from Radcliffe et al (2006) in the following categories: differentially expressed in same direction at $p<0.05$ (blue, 195 genes), at $p<0.1$ (red, 32 genes) and $p>0.1$ (green, 43 genes). Twenty seven genes had opposite relative expression values between datasets (purple).

Table 2.2 Over-representation analyses for DEGs in whole brain and striatum.

Brain Region	General category	Classification term[a]	Resource[b]	# genes[c]	p-value[d]	Corrected p-value[e]
Striatum	Synapse/Signaling	Potassium channel activity	Gene Ontology	14	9.92E-08	1.01E-05
		G-protein coupled receptor signaling pathway	Gene Ontology	23	3.01E-06	5.00E-04
		Signal transduction	Gene Ontology	68	1.00E-03	2.00E-03
		Neuron development	Gene Ontology	17	7.15E-05	2.00E-03
		Non-odorant G-protein coupled receptors	Wikipathways	13	2.00E-04	3.40E-03
		Calcium signaling pathway	KEGG	12	2.00E-04	5.60E-03
		Dopamine receptor activity	Gene Ontology	2	9.00E-04	6.10E-03
		Beta-adrenergic receptor kinase activity	Gene Ontology	2	9.00E-04	6.10E-03
		Neuroactive ligand-receptor interaction	KEGG	11	4.00E-04	7.50E-03
		Negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	Gene Ontology	4	5.00E-04	8.80E-03
		Synapse	Gene Ontology	15	4.20E-03	3.65E-02
		Opioid peptide activity	Gene Ontology	2	8.40E-03	4.18E-02
		Gap junction	KEGG	7	4.30E-03	4.82E-02
	Behavior	Response to amphetamine	Gene Ontology	4	1.00E-04	2.00E-03
	Ribosome	Cytoplasmic ribosomal proteins	Wikipathways	13	1.22E-07	4.15E-06
		Ribosome	KEGG	13	2.61E-07	1.46E-05
		Ribosome	Gene Ontology	13	1.10E-03	1.41E-02
	Cell types	Neuron	Cahoy et al	96	1.00E-16	
		Oligodendrocyte	Cahoy et al	26	2.50E-03	
	Lore QTL	*Lore4* Chr11:79000000-108000000	Bennett et al	13	4.10E-02	

Table 2.2 continued. Over-representation analyses for DEGs in whole brain and striatum.

Brain Region	General category	Classification term[a]	Resource[b]	# genes[c]	p-value[d]	Corrected p-value[e]
Whole Brain	Ribosome	Cytoplasmic ribosomal proteins	Wikipathways	6	1.04E-05	7.28E-05
		Ribonucleoprotein complex	Gene Ontology	8	4.20E-03	2.85E-02
		Ribosome	KEGG	7	9.19E-07	9.19E-06
	Cell membrane	Extracellular region	Gene Ontology	13	7.00E-04	1.86E-02
	Metabolic pathway	Retinol metabolism	Wikipathways	2	1.17E-02	4.10E-02
	Immune	MHC protein complex	Gene Ontology	2	3.80E-03	2.85E-02
	Cell types	Astrocyte	Cahoy et al	10	2.60E-02	
	Lore QTL	*LoreChr3* Chr3:130000000-155000000	Bennett et al	2	2.70E-02	

[a]Term used to classify related genes.
[b]Resource used for classification, Gene Ontology, KEGG, NCBI Entrez Gene, Wikipathways, Cahoy et al (2008), or Bennett et al (2006)/personal communication with Dr. Bennett.
[c]Number of differentially expressed genes in each category.
[d]Uncorrected hypergeometric p-value testing whether number of DEGs in each term more than expected.
[e]Benjamini-Hochberg corrected p-values.

Additionally, *LoreChr3* on chromosome 3 was enriched with WB DEGs (2 genes, hypergeometric p=0.027). In striatum, *Lore4* on chromosome 11 (13 genes, hypergeometric p=0.041) was enriched (Table 2). The set of ST DEGs was also enriched for genes previously shown to be at least 3-fold over-expressed in oligodendrocytes (26 genes, hypergeometric p=0.0025) and neurons (96 genes, hypergeometric $p<1x10^{-16}$). The set of WB DEGs was enriched for astrocyte-related genes (10 genes, hypergeometric p=0.026).

2.4.5 Weighted gene co-expression network analysis (WGCNA)

A single WGCNA of all 12 samples produced 16 distinct clusters (modules) of similarly expressed genes. The number of genes in each module ranged from 24 to 8,288. Each gene was assigned to a colored module, and no grey module (representing non co-expressed genes) was created (Figure 2.3). Module robustness was tested using two methods. First, in each module, permutation testing confirmed that average module adjacency was always greater than the mean of 1000 randomly sampled "modules" of equal size (all modules p<0.001). Second, all modules were shown to display higher scaled intramodular connectivity compared to scaled extramodular connectivity.

2.4.6 WGCNA gene modules enriched with differentially expressed genes

To determine whether each module contained more differentially expressed genes than expected, the number of observed differentially expressed genes in each module was compared to the hypergeometric distribution of the expected number of differentially

expressed genes. Six modules were enriched with striatum DEGs (blue, cyan, green, greenyellow, magenta, and yellow) (Table 2.3). Of the 336 striatal DEGs, 96 out of 3211 in the blue module were differentially expressed (hypergeometric p=0.025), 8 of 76 in the cyan module (hypergeometric $p=3.67x10^{-4}$), 12 of 123 in the green module (hypergeometric $p=1.48x10^{-4}$), 12 of 171 in the greenyellow module (hypergeometric $p=9.1x10^{-4}$), 9 of 87 in the magenta module (hypergeometric $p=2.59x10^{-4}$), and 19 of 299 in the yellow module (hypergeometric $p=2.59x10^{-4}$). Four modules were enriched with whole brain DEGs (darkred, green, magenta, and yellow, Table 3). Of the 90 whole brain DEGs, 1 of 24 in darkred were differentially expressed (hypergeometric p=0.042), 7 of 123 in green (hypergeometric $p=7.1x10^{-6}$), 10 of 87 in magenta (hypergeometric $p=1.45x10^{-10}$), and 12 of 299 in yellow (hypergeometric $p=5.27x10^{-7}$). All p-values have been adjusted for multiple corrections according to the Benjamini-Hochberg method, using the p.adjust function in R.

2.4.7 Module eigengenes associated with strain/region differences

We calculated the 1st principle component (PC) of each module using the moduleEigengenes command from the WGCNA R-package. The 1st PC, or module eigengene, represents the sample-specific expression levels if each module were reduced to a single gene (Hierarchical clustering of module eigengenes is shown in Figure 2.4). An analysis of variance (ANOVA) of the module eigengenes (Figure 2.5) resulted in strain differences in four modules: green ($F_{1,8}=274.6$, $p=1.78x10^{-7}$), grey60 ($F_{1,8}=46.11$, $p=1.39x10^{-4}$), magenta ($F_{1,8}=258.3$, $p=2.26x10^{-7}$) and yellow ($F_{1,8}=65.06$, $p=4.12x10^{-5}$). Three modules were different by region—black ($F_{1,8}=78.03$, $p=2.13x10^{-5}$),

Figure 2.3 Hierarchical clustering and dynamic tree cut.

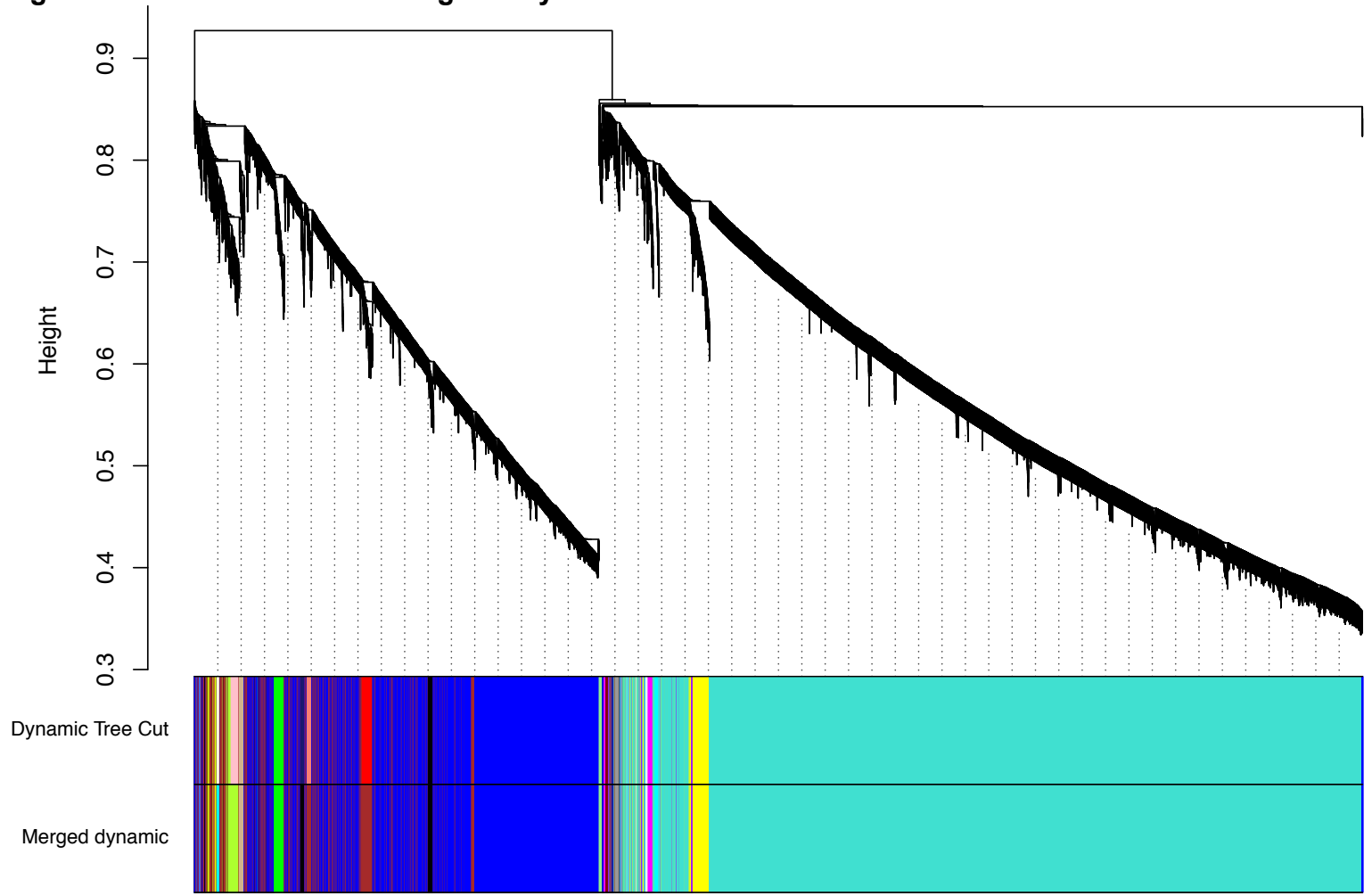

Figure 2.3 shows the results of the hierarchical clustering algorithm and the dynamic tree cut. The y-axis represents a dissimilarity measurement based on topological overlap, with the more similar topological overlaps corresponding to lower heights. Each branch of the dendrogram represents one gene. Branches of the dendrogram are “pruned” into modules, corresponding to each color in the bottom rows. The top color row shows the module grouping after the initial dynamic tree cut (21 modules), while the bottom color row shows the module grouping after merging similar modules (16 remaining modules).

Table 2.3. WGCNA Co-expression module characteristics.

Module[a]	#genes	Module eigengene significant[b]	DEG enrichment[c]	Cell type enrichment[d]	Top genes[e]	*Lore* QTL[f]	DEG[g]
black	165	Region (2.13E-05)			*Matk* *Kcnh3* *Psd* *Tmem191c* *Ppp2r2c*		
yellow	299	Strain (4.12E-04)	ST (2.60E-04) WB (5.26E-07)		*Ptprn* *Eif3k* *Glt25d1* *Tnip1* *Tpd52*	Lore1	
brown	1082	Region (1.11E-03)		Oligodendrocyte (2.31E-07)	*Rps6ka4* *Gsn* *Rbx1* *Ephb1* *Icam5*		
cyan	76		ST (3.66E-04)		*Robo3* *Kalrn* *Kcns1* *Cacnb3* *Sytl2*	Lore2b	ST ST
greenyellow	171	Strain (6.63E-06) Region (2.37E-04)	ST (9.10E-04)		*6030458C11Rik* *Selplg* *4933439F18Rik* *4632428N05Rik* *Gm10116*		ST ST, WB ST
magenta	87	Strain (2.26E-07)	ST (2.60E-04) WB (1.45E-10)		*Gm10516* *Folh1* *Prss50* *Rnasel* *4930452B06Rik*		ST, WB WB WB ST, WB
grey60	36	Strain (1.39E-04)			*2610002J02Rik* *Polr1b* *Lama2* *Chi3l1* *Adi1*	Lore2a	
darkred	24		WB (4.23E-02)		*Rnd2* *Uchl1* *Add1* *Ncan* *Tstd2*	Lore4	

Table 2.3 continued. WGCNA Co-expression module characteristics.

Module[a]	#genes	Module eigengene significant[b]	DEG enrichment[c]	Cell type enrichment[d]	Top genes[e]	*Lore* QTL[f]	DEG[g]
green	123	Strain (1.78E-07)	ST (1.48E-04) WB (7.10E-06)		*Tmem181a* *A530054K11Rik* *Copb1* *Tmem181b-ps* *Trmt6*	Lore2a	ST, WB ST ST
blue	3211	Strain (2.20E-03) Region (2.01E-06)	ST (2.51E-02)		*Rasgrp1* *Ppp1r9a* *Pde7b* *Nexn* *Rgs4*	LoreChr3	ST ST ST
turquoise	8288	Region (1.63E-06)		Neuron (5.45E-04) Astrocyte (2.07E-02)	*Gm672* *Kndc1* *Pcdh1* *Plxna1* *Slc20a2*		ST ST

[a]Gene co-expression module produced by WGCNA. Modules not significant for eigengene difference, DEG or cell-type enrichment are not shown.
[b]Results of Analysis of Variance of the first principle component, or module eigengene, of expression values for each module. Each module eigengene was tested across strain and region.
[c]Module over-representation of DEGs from either ST or WB.
[d]Module over-representation of genes expressed at least 3-fold higher in specific cell types, neurons, astrocytes, or oligodendrocytes (Cahoy et al, 2008).
[e]The top five most inter-connected genes in each module.
[f]*Lore* QTL region containing corresponding top gene.
[g]Region of differential expression of corresponding hub gene.

Figure 2.4 Hierarchical clustering of module eigengenes.

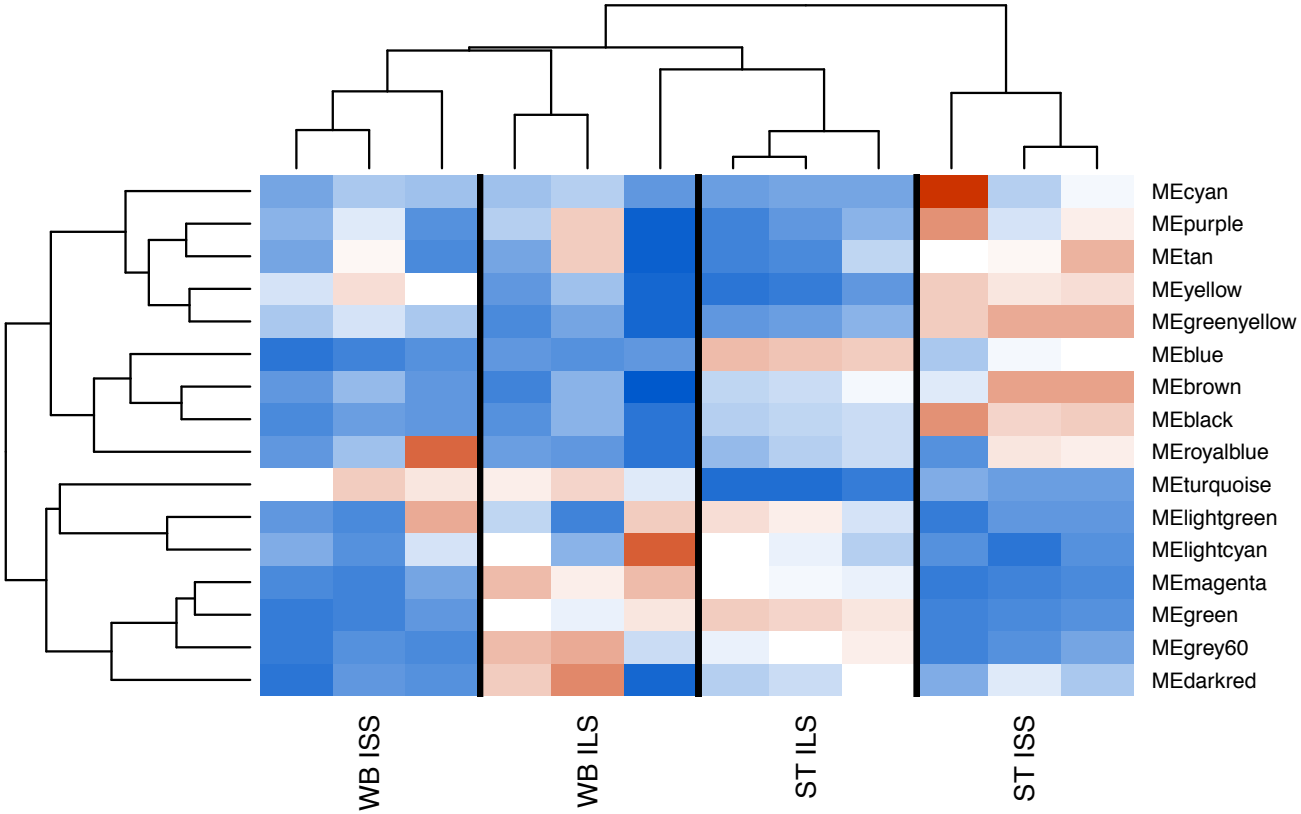

Figure 2.4 shows the hierarchical cluster of each module eigengene (rows) and each sample (columns). Eigengene values range from approximately -0.5 to 0.9, representing sample specific expression levels. Higher expression is denoted with red colors, lower expression by blue colors.

Figure 2.5 WGCNA module eigengene expression levels.

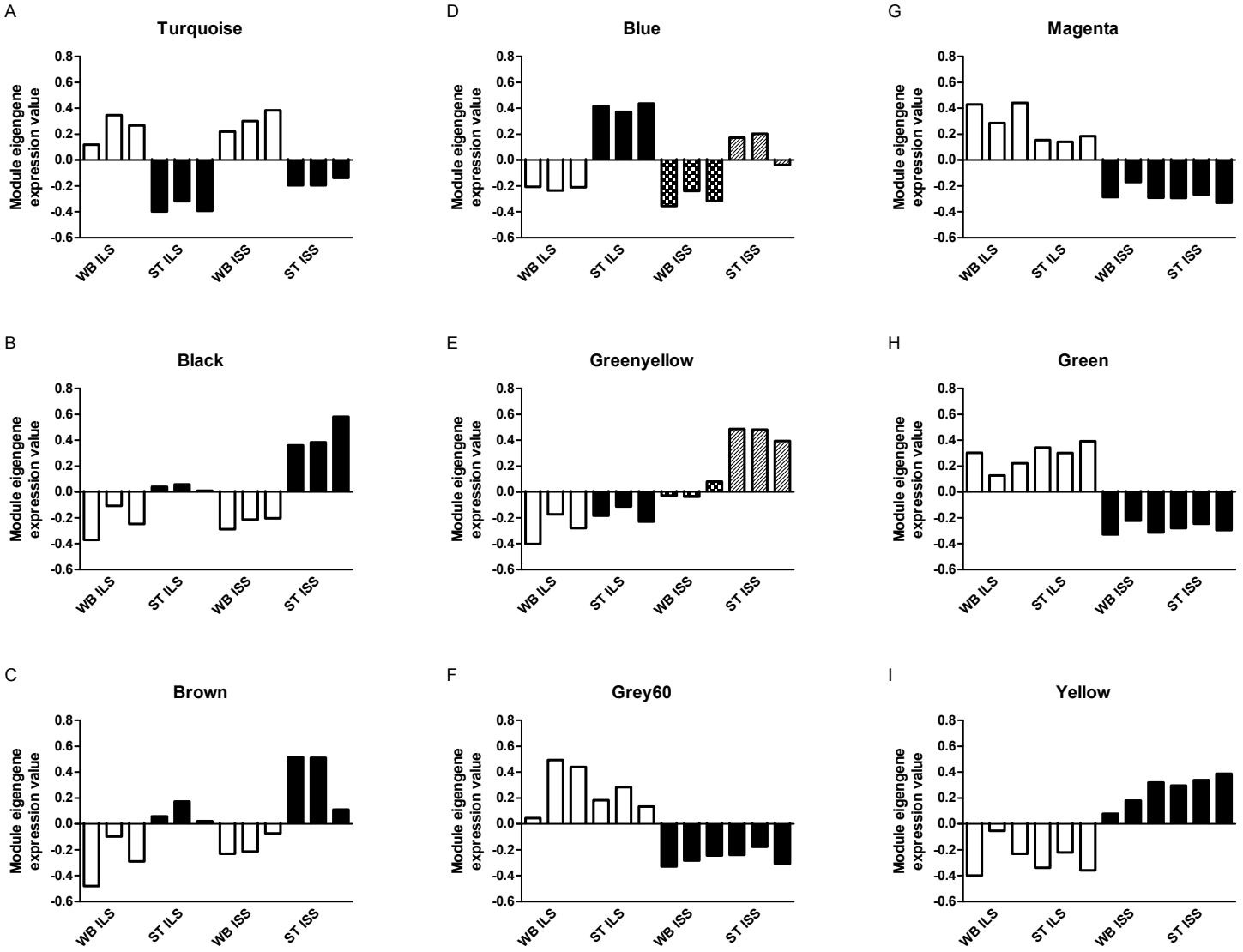

Figure 2.5 displays the calculated expression level of module eigengenes, the first principle component of each

module expression pattern. Individual mouse samples (bars) are in groups of 3 for each set of whole brain ILS, striatum ILS, whole brain ISS and striatum ISS. Only module eigengenes significant for strain or region differences are shown. Module eigengenes reduce the expression value of all genes in the module to one value per sample. An ANOVA of each module eigengene reveals modules different by region (A-C), both region and strain (D-E), and strain (F-I). No module eigengenes had significant strain x region interactions

brown ($F_{1,8}$=24.62, p=1.11x10^{-3}), and turquoise ($F_{1,8}$=154.6, p=1.63x10^{-6}). Two modules were different for both strain and region: blue ($F_{1,8}$=19.61, p=0.0022, strain; $F_{1,8}$=146.43, p=2.01x10^{-6}, region) and greenyellow ($F_{1,8}$=106.8, p=6.63x10^{-6}, strain; $F_{1,8}$=39.49, p=2.37x10^{-3}, region), No module eigengenes had significant strain x region interaction effects. P-values were considered significant when less than 0.05/16=0.003125.

2.4.8 Cell type over-representation in WGCNA modules

Using genes identified as being significantly over-expressed, by at least 3-fold, in neurons, astrocytes, or oligodendrocytes, we tested whether modules were enriched for these sets of genes (Table 3 and Figure 2) [112]. Of the 13,802 genes used in the WGCNA, 1,099 (neuron), 803 (astrocyte), and 556 (oligodendrocyte) had been identified as being over-expressed by at least 3-fold in each cell type. The turquoise module was enriched with 721 neuron genes (hypergeometric p=5.45x10^{-4}) and 522 astrocyte genes (hypergeometric p=0.021). The brown module was enriched with 81 oligodendrocyte genes (hypergeometric p=2.31x10^{-7}). All p-values were adjusted for the Benjamini-Hochberg false discovery rate.

2.4.9 Gene module hub gene identification

WGCNA identifies networks of interconnected genes, and it is possible to further identify the most interconnected genes in each module. The top five most interconnected genes (hub genes) in the eleven modules either enriched for DEGs or different across strain or region are listed in Table 3. Seventeen DEGs were identified

as hub genes, 11 ST DEGs, 2 WB DEGs, and 4 DEGs from both ST and WB. Nine genes located within *Lore* QTLs were also hub genes. Of the six modules identified as different across strain, four had DEGs as hub genes. In the blue module, phosphodiesterase 7b (*Pde7b*), nexilin, F-actin binding protein (*Nexn*), and regulator of G-protein signaling 4 (*Rgs4*), all ST DEGs, are hub genes. Three ST DEGs in the greenyellow module were hub genes, *6030458C11Rik*, *4933439F18Rik*, and selectin P ligand (*Selplg*). Additionally, three genes in the green module, *A530054K11Rik*, coatamer protein complex subunit beta 1 (*Copb1*), and transmembrane protein 181b pseudogene (*Tmem181b-ps*), along with four genes in the magenta module, *Gm10516*, folate hydrolase (*Folh1*), protease, serine 50 (*Prss50*), and ribonuclease A, family 1 (*Rnasel*), are differentially expressed in either ST, WB, or both.

2.4.10 Functional group over-representation in WGCNA modules

Co-expression modules were analyzed using WebGestalt to test for functional group over-representation. In the six modules modules differing by strain, several signaling pathways were over-represented, including mitogen-activated protein kinase (MAPK) signaling (blue, yellow), peroxisome proliferator activated protein (PPAR) signaling (blue, greenyellow), transforming growth factor (TGF) beta signaling (blue, greenyellow), nuclear factor κB (NF-κB) signaling (blue, magenta, yellow), and toll-like receptor (TLR) signaling (blue, yellow). Genes involved in regulating the actin cytoskeleton were enriched in blue, green, and yellow. Complement and coagulation cascades were enriched in the magenta module. All group p-values range from $1.1 \times 10^{-}$

[38] - 0.048 and have been corrected for multiple testing and were significant at <5% false discovery rate.

2.4.11 Identification of cis-regulated Lore QTL genes

Utilizing publicly accessible databases of recombinant inbred gene expression data from the online WebQTL tool (www.genenetwork.org), we identified differentially expressed genes from both striatum and whole brain, as well as hub genes, in *Lore* QTL regions that have evidence of *cis*-regulation. Each hub gene and DEG lying in *Lore* QTL regions was interrogated. A total of 11 genes showed evidence of *cis*-regulation. Three DEGs, alanine-glyoxylate aminotransferase 2-like 1 (*Agxt2l1*) located in *LoreChr3*, ras association (RalGDS/AF-6) domain family member 2 (*Rassf2*) located in *Lore2a* and keratin 12 (*Krt12*) located in *Lore4* were differentially expressed in both WB and ST, and show strong evidence of *cis*-regulation. Six genes differentially expressed in the ST, *Lore1* genes regulated endocrine-specific protein 18 (*Resp18*) and serine peptidase inhibitor, clade E, member 2 (*Serpine2*), *Lore3* gene centromere protein t (*Cenpt*), *Lore4* genes Rap guanine nucleotide exchange factor GEF-like 1 (*Rapgefl1*), myosin light chain 4 (*Myl4*), and keratin 9 (*Krt9*), *Lore5* all show evidence of *cis*-regulation. The WB DEG and *LoreChr3* gene DNA-damage-inducible transcript 4-like (*Ddit4l*), as well as the grey60 module hub gene, polymerase (RNA) I polypeptide B (*Polr1b*) also could be *cis*-regulated.

2.4.12 Sequence differences

Of the *Lore* QTL genes with evidence of *cis*-regulation, only *Resp18* and *Agxt2l1* did not have any detectable sequence differences (Table 2.4). Of note, an unnamed missense single nucleotide polymorphism (SNP) in *Serpine2*, resulting in an isoleucine to valine substitution (I313V) in both ILS and ISS mice was observed. Four missense SNPs in *Cenpt*, three of which were unnamed were only observed in ISS. More unnamed SNPs were observed in *Myl4, Polr1b*, and *Ddit4l*. Also notable are the multitude of polymorphisms in 3' UTR of *Rassf2*. According to www.microrna.org, these polymorphisms could potentially disrupt the binding sites of multiple miRNAs.

Fifteen genes previously reported to contain coding sequence differences were examined, and each polymorphism was confirmed in twelve of the genes [93]. Low expression levels in *Tgfb1* and *Pth2r* (named *Pthr* in original paper) made it impossible to identify polymorphisms. *Znf133* has since been classified as a pseudogene, although it is expressed in our sample, and several single nucleotide polymorphisms are confirmed; however, frame shift mutations could not be confirmed. Although there are numerous sequence differences between the two strains, complete identification and classification of polymorphisms was beyond the scope of the study.

Table 2.4 Polymorphisms in probable cis-regulated *Lore* QTL DEGs.

Gene[a]	Polymorphism[b]	Locus[c]	Feature[d]	Type[e]	Strain[f]
	rs13469719	Chr1:79,790,995	3' UTR	G/T	ILS
	unnamed	Chr1:79,798,079	Exon 6	missense I313V	Both
Serpine2	rs32034294	Chr1:79,807,313	Exon 4	A/G synonymous	Both
	rs49368455	Chr1:79,813,448	Exon 3	C/T synonymous	Both
	rs13469718	Chr1:79,855,118	5' UTR	A/G	Both
	rs29426703	Chr11:98,712,635	3' UTR	T/C	ISS
Rapgefl1	rs27026239	Chr11:98,713,562	3' UTR	T/C	ISS
	rs27026233	Chr11:98,714,256	3' UTR	A/G	ISS
	rs27275027	Chr2:131,818,710	3' UTR	A/C	ILS
	rs47809900	Chr2:131,818,753	3' UTR	A/G	ILS
	unnamed	Chr2:131,818,755	3' UTR	G/A	ILS
	unnamed	Chr2:131,818,837	3' UTR	C/A	ILS
	unnamed	Chr2:131,818,839	3' UTR	C/T	ILS
	unnamed	Chr2:131,818,897	3' UTR	G/T	ILS
	unnamed	Chr2:131,818,901	3' UTR	G/A	ILS
	unnamed	Chr2:131,818,923	3' UTR	T/C	ILS
	unnamed	Chr2:131,819,139	3' UTR	C/T	ILS
	unnamed	Chr2:131,819,142	3' UTR	A/G	ILS
	rs27275025	Chr2:131,819,202	3' UTR	T/C	ILS
	rs27275024	Chr2:131,819,243	3' UTR	C/G	ILS
	unnamed	Chr2:131,819,252	3' UTR	T/C	ILS
	rs27275023	Chr2:131,819,380	3' UTR	G/A	ILS
	unnamed	Chr2:131,819,562	3' UTR	G/A	ILS
	rs27275021	Chr2:131,819,638	3' UTR	A/G	ILS
	rs27275020	Chr2:131,819,688	3' UTR	T/C	ILS
	rs27275019	Chr2:131,819,740	3' UTR	G/C	ILS
	rs27275018	Chr2:131,819,812	3' UTR	A/G	ILS
	rs27275017	Chr2:131,819,924	3' UTR	G/A	ILS
	unnamed	Chr2:131,819,969	3' UTR	T/C	ILS
	rs27275016	Chr2:131,820,136	3' UTR	C/T	ILS
Rassf2	rs27275015	Chr2:131,820,205	3' UTR	G/A	ILS
	rs27275014	Chr2:131,820,301	3' UTR	G/A	ILS
	unnamed	Chr2:131,820,730	3' UTR	T/C	ILS
	unnamed	Chr2:131,820,740	3' UTR	T/C	ILS
	rs27275012	Chr2:131,820,772	3' UTR	A/G	ILS
	rs27275011	Chr2:131,820,921	3' UTR	G/A	ILS
	rs27275010	Chr2:131,820,989	3' UTR	G/A	ILS
	unnamed	Chr2:131,821,093	3' UTR	A/G	ILS
	rs27275008	Chr2:131,821,250	3' UTR	A/G	ILS
	unnamed	Chr2:131,821,376	3' UTR	G/A	ILS
	rs27275004	Chr2:131,821,441	3' UTR	A/G	ILS
	rs27275002	Chr2:131,821,504	3' UTR	G/A	ILS
	unnamed	Chr2:131,821,555	3' UTR	T/C	ILS
	rs27275001	Chr2:131,821,600	3' UTR	T/C	ILS
	unnamed	Chr2:131,821,742	3' UTR	T/G	ILS
	unnamed	Chr2:131,821,745	3' UTR	T/C	ILS
	unnamed	Chr2:131,821,949	3' UTR	A/G	ILS
	unnamed	Chr2:131,822,010	3' UTR	A/G	ILS
	rs27274999	Chr2:131,822,021	3' UTR	A/G	ILS
	unnamed	Chr2:131,822,024	3' UTR	G/A	ILS
	unnamed	Chr2:131,822,028	3' UTR	G/A	ILS
	rs27274997	Chr2:131,822,058	3' UTR	A/C	ILS
	rs27274996	Chr2:131,822,132	Exon 11	G/A synonymous	ILS
	rs27274994	Chr2:131,822,183	Exon 11	A/G synonymous	ILS

Table 2.4 continued. Polymorphisms in probable cis-regulated *Lore* QTL DEGs.

Gene	Polymorphism	Locus	Feature	Type	Strain
Myl4	rs29426930	Chr11:104,438,799	Exon 1	missense T12A	ILS
	unnamed	Chr11:104,445,898	Exon 6	A/C synonymous	ILS
	unnamed	Chr11:104,445,919	Exon 6	T/C synonymous	ILS
Krt12	rs51628282	Chr11:99277014	3' UTR	T/C	ISS
	rs27088547	Chr11:99277037	3' UTR	G/A	ISS
	rs27088526	Chr11:99278249	Exon 7	A/G synonymous	ISS
	rs27088536	Chr11:99277370	Exon 8	G/A synonymous	ISS
Krt9	rs27088361	Chr11:100052809	Exon 2	G/A synonymous	ILS
	rs27088362	Chr11:100052788	Exon 2	G/A synonymous	ILS
	rs52613970	Chr11:100049988	Exon 7	missense Y631H	ILS
Polr1b	rs45674576	Chr2:128928830	Exon 2	C/T synonymous	ISS
	unnamed	Chr2:128939427	Exon 9	T/G synonymous	ISS
	rs27448743	Chr2:128944898	Exon 12	C/T synonymous	ISS
	rs27448705	Chr2:128951629	Exon 15	missense M1069V	ISS
	rs27448701	Chr2:128951933	3' UTR	C/T	ISS
Ddit4l	unnamed	Chr3:137287209	Exon 2	T/G synonymous	ILS
	rs50093517	Chr3:137290001	3' UTR	C/T	ISS
	rs48364418	Chr3:137290393	3' UTR	G/A	Both
	unnamed	Chr3:137290499	3' UTR	C/T	ILS
	rs51973625	Chr3:137290781	3' UTR	C/T	ISS
	rs46955320	Chr3:137290842	3' UTR	A/G	Both
	rs50360881	Chr3:137290953	3' UTR	T/C	Both
	rs31235381	Chr3:137290990	3' UTR	C/G	Both
	rs30309919	Chr3:137291052	3' UTR	A/T	Both
	rs30112060	Chr3:137291124	3' UTR	T/A	Both
	rs31345253	Chr3:137291128	3' UTR	T/C	Both
	rs31048748	Chr3:137291212	3' UTR	T/A	Both
Cenpt	unnamed	Chr8:108375915	5' UTR	G/A	ISS
	unnamed	Chr8:108372676	Exon 7	missense D232E	ISS
	unnamed	Chr8:108370923	Exon 8	missense M292V	ISS
	unnamed	Chr8:108370885	Exon 8	C/G synonymous	ISS
	unnamed	Chr8:108369315	Exon 11	missense E388D	ISS
	unnamed	Chr8:108369303	Exon 11	A/G synonymous	ISS
	rs48755141	Chr8:108369029	Exon 12	missense S457A	ISS

[a]Gene in *Lore* QTL region with evidence of *cis*-regulation.
[b]dbSNP ID if previously annotated.
[c]Chromosome and base position based on mouse genome build 9 (ensembl.org).
[d]Gene feature where the polymorphism is found (intron regions were not included due to low coverage as a consequence of poly-A enrichment).
[e]Type of polymorphism/resulting amino acid substitution.
[f]Strain that is different from the reference genome.

2.5 Discussion

Loss of righting reflex in response to acute ethanol has been well studied in the ILS and ISS strains, and respective QTLs have been identified and replicated using recombinant panels, both LSxSS and LxS [52-56, 72, 92]. The goal of this study was to identify baseline differences in gene expression and co-expression between these two selected inbred strains, which will provide insight into the underlying biology that contributes to their differential sensitivity to ethanol. While previous studies have identified candidate genes based on expression differences, this study uses multiple methods, differential expression, Weighted Gene Co-expression Network Analysis, identification of *cis*-regulated *Lore* QTL genes, and identification of sequence differences in coding and un-translated regions. The use of RNA-Seq technology, as opposed to previous use of microarray, provides higher dynamic range, lower background noise, improved network characteristics, and the elimination of hybridization issues due to polymorphisms and annotation [86-88]. In this study, 90 genes in WB and 336 in ST samples were differentially expressed. We prioritize genes that are located in previously identified Lore QTL regions for future study. Eight WB and 31 ST DEGs are located in *Lore* QTL regions. While the total number of QTL genes is no different than chance, two *Lore* QTL regions were enriched for DEGs, *LoreChr3* on chromosome 3 was enriched with WB DEGs and *Lore4* on chromosome 11 was enriched for ST DEGs. This could potentially signify regional differences in gene expression, and future transcriptome examinations may identify regions enriched with other *Lore* QTL genes.

Two previously identified candidate genes [72], *Rassf2 (*ras association (RalGDS/AF-6) domain family member 2), located in *Lore2a*, and *Myo1d (Lore4* gene

myosin 1d) were identified by our analysis as differentially expressed in both ST and WB. MacLaren sequenced the promoter region of *Rassf2,* finding several polymorphisms [57]. One advantage of RNA-Seq is the acquisition of the genetic sequence of exons and untranslated regions (UTRs). Examination of the 3' UTR of *Rassf2* shows distinct genotypes. ISS mice have the C57Bl/6J haplotype, while the ILS 3' UTR shows many SNPs, several unnamed in dbSNP. Since the 3' UTR is implicated in post-transcriptional regulation, including microRNA binding sites, the polymorphisms could account for some of the previously observed differences in expression. The observed ILS polymorphisms disrupt the consensus sequences for binding sites of 9 miRNAs (www.microrna.org). We were unable to detect expression levels for these miRNAs, so whether they affect expression levels of *Rassf2* remains to be seen. We present evidence that several genes, including *Rassf2*, are *cis*-regulated, meaning that polymorphisms in gene regions between the two strains could contribute to differences in gene expression. If these are *cis*-regulated, it is likely that differences in gene expression could be explained by genetic polymorphisms in either coding regions or UTRs. Furthermore, while synonymous polymorphisms in exons may not affect protein function, they are indicative of distinct haplotypes between strains and of possible polymorphisms in intergenic or intronic regions that could affect expression. It is not clear how *Rassf2* and *Myo1d* could influence ethanol-related behavior. *Rassf2* has been characterized as a pro-apoptotic gene, residing in the nucleus and binding K-Ras, inducing apoptosis [115]. Differences could also arise from the role of *Myo1d* in the development of the nervous system [116]. Taken together, it is possible that strain specific neural development could lead to phenotypic differences.

Located in *Lore2a* is the DEG prodynorphin, *Pdyn*. More highly expressed in ST of ILS mice, *Pdyn* is differentially expressed in other animal models of ethanol behaviors. Consistent with our findings, low drinking ANA rats have increased levels of striatal *Pdyn* compared to higher drinking AA rats [117]. Another opioid precursor gene, proenkephalin, *Penk*, is also more highly expressed in the ST of ILS mice. While the difference between strains in opioid signaling has not been explored in depth, it has been shown that SS and LS mice differ in response to morphine injection and withdrawal [118]. Another QTL gene, in *Lore4*, *Ppp1r1b*, which codes for protein phosphatase 1 regulatory unit 1b, also known as DARPP-32, has been implicated in the neurobiological response to many drugs of abuse [119]. *Ppp1r1b* is expressed in striatal medium spiny neurons (MSNs), and plays a large role in the cellular response to dopaminergic signaling.

In addition to genes from *Lore* QTL regions, transthyretin (*Ttr*) on chromosome 19 was also identified in both samples as being differentially expressed. Gamma-protein kinase C (PKC-γ) null mutant mice and their wild-types have similar ethanol-related behaviors as the ISS and ILS mice, and these differences were correlated with baseline *Ttr* expression, which is higher in mutant mice [44]. Similar to the ISS mice, PKC-γ null mutants are less sensitive to acute ethanol than their wild-type littermates [120], and voluntarily consume more ethanol [121]. Likewise, baseline expression of *Ttr* in ISS mice is higher relative to ILS mice. While it is unknown whether a chronic ethanol diet would increase expression of *Ttr* in the ISS mice, as in the PKC-γ null mutants, future confirmation would further implicate *Ttr* in ethanol behavior. Also of interest are the 14 potassium channel subunits differentially expressed in the striatum; as potassium

channels have been implicated in responses to ethanol [122-125] and the cumulative effect of differential expression of all of these channels could contribute to the difference in ethanol sensitivity between the strains.

While RNA-Seq is thought to offer several advantages over microarrays, it still suffers a problem inherent to any massively parallel method: finding the appropriate statistical balance between type I and type II errors. Validation by an independent method is one approach and here we have used microarray data to validate the RNA-Seq DEGs. The results are similar, perhaps slightly better, to a comprehensive comparison of RNA-Seq to hybridization microarrays conducted by Bottomly *et al.* (2011); i.e., they found that 48.4% of genotype-dependent RNA-Seq DEG were also DE on the Affymetrix platform and we found that this was true for 65.7% of our RNA-Seq DEG, although our statistical criteria was somewhat less stringent. In addition to the possibility of statistical errors, reasons for less than perfect consistency between RNA-Seq and microarrays probably include the broader dynamic range of RNA-Seq and, more importantly, the likelihood of genotype effects on transcript isoform abundance meaning for microarrays, quantification of a given transcript is dependent on probeset location [126]. Indeed, we have seen hints of evidence for strain-by-isoform interactions for some of the microarray probesets that were not significant, although this particular RNA-Seq dataset is not ideal for a comprehensive splice variant analysis.

Using WebGestalt to identify over-represented groups in our sets of DEGs, we identified several distinct groups of differentially regulated gene systems. In ST, there were many DEGs involved in signal transduction and synaptic signaling. In addition to functional groups, we identified cell type specific (neurons, astrocytes, and

oligodendrocytes) genes over-represented in each set of DEGs. The set of ST DEGs was enriched for neuron and oligodendrocyte genes. Specifically, the set of 127 DEGs up-regulated in ILS mice was only enriched for neuronal genes, while the set of 209 DEGs up-regulated in ISS mice was enriched for all three types of cells. This suggests that while there are differences in neuronal processes between the two strains, there may be more important differences in glial related processes. This holds up when looking at WB DEGs, as the set of WB DEGs is enriched only for astrocyte related genes.

To further characterize strain specific differences in gene expression, we employed the agnostic network analysis tool WGCNA, which clustered genes based on topological overlap dissimilarity. The results of the WGCNA display its usefulness at analyzing large expression datasets. Gene modules were enriched for cell specific genes, and module eigengenes highlight strain- and region-specific differences. However, there is a limitation on the interpretations due to the small sample size in our study, even though each module passed strict robustness testing. No hub genes were immediately identifiable as strong candidate genes, however it is important to acknowledge that the WGCNA identifies networks of related genes, and the effect of any single gene could be minimal. It differs in this way from the differential expression analysis, where the genes with the largest differences in expression, and possibly having larger effects, are identified. In this analysis, we were less confident in some of the smaller modules where some samples appeared to be outliers, but more confident of modules showing consistent expression levels within groups (either regional or

strain). These patterns of expression are striking, and show that genes can be consistently co-expressed at different levels depending on region or strain.

Of the 16 gene co-expression networks (modules), three were enriched for ST DEGs, one for WB DEGs, and three were enriched for both ST and WB DEGs. This made it possible to identify not only DEGs, but also gene networks in which those DEGs reside. Functional group over-representation of DEG-enriched modules revealed many genes related to neuronal structure and function, as well as transcriptional regulation. Interestingly, these modules were enriched for several signaling pathways, including MAP Kinase signaling pathways, previously shown to regulate ethanol behaviors [127].

One module, turquoise, was enriched with neuron genes. Since this module eigengene differed across region, and not strain, this module is most likely composed of neuronal genes differentially expressed due to regional differences, and given that this is the largest module, most of the co-expression differences can likely be due to brain regional differences. Of the six modules different across strain, five were enriched for ST DEGs, while three of those were also enriched for WB DEGs.

Utilizing RNA-Seq technology to identify gene expression differences and gene co-expression networks has provided insight into the differences between ILS and ISS mice. Genes previously identified as candidates from expression/QTL studies, *Rassf2*, *Myo1d*, and drug response studies, *Pdyn, Penk, Ppp1r1b*, and *Ttr* are again implicated. While these differences exist, this study is not designed to specify causal differences. Therefore, it is important for future research to focus on manipulation, genetic or pharmacological, of genes and gene networks to further elucidate the differences between these strains, in order to understand the cause of ethanol-related behaviors.

Chapter 3

Mesolimbic transcriptional response to hedonic substitution of voluntary exercise and voluntary ethanol consumption

Todd M Darlington and Marissa A Ehringer

Submitted to *Behavioral Brain Research*.

3.1 Abstract

The mesolimbic dopaminergic pathway has been implicated in many rewarding behaviors, including the consumption of ethanol and voluntary exercise. It has become apparent that different rewarding stimuli activate this pathway, and therefore it is possible for these behaviors to influence each other, i.e. hedonic substitution. Using adult female C57BL/6Ibg mice, we demonstrate that voluntary access to a running wheel substantially reduces the consumption and preference of ethanol. Furthermore, we examined gene expression of several genes involved in regulating the mesolimbic dopaminergic pathway, which we hypothesized to be the focal point of hedonic substitution. In the striatum, we observed a reduction in mRNA expression of *Drd1a* due to exercise. Hippocampal *Bdnf* mRNA increased in response to exercise and decreased in response to ethanol. Furthermore, there was an interaction effect of exercise and ethanol on the expression of *Slc18a2* in the midbrain. These data suggest an important role for this pathway, and especially for *Bdnf* and *Slc18a2* in regulating hedonic substitution.

3.2 Introduction

Abuse of alcohol is a leading cause of preventable disease and death worldwide, affecting an estimated 76.3 million people [1]. Extensive research is being conducted on the development of alcohol use disorders, and a number of candidate genes have shown association with alcohol use [128]. Ethanol interacts with a variety of subcellular components comprising many of the known neurotransmitter systems including the mesolimbic dopaminergic pathway [129-131]. It has been proposed that a common pathway exists for addiction, and cross-tolerance between drugs of abuse, as well as co-abuse has been observed [131].

McMillan (1978) first reported the behavioral interaction of exercise and ethanol [59, 132]. Since then, several groups have shown that access to exercise can influence voluntary ethanol intake [60-62]. Recent work in our laboratory supported the hypothesis that wheel-running may influence the reinforcing effects of ethanol [63]. This concept of hedonic substitution has been implemented in exercise intervention programs for humans consuming high quantities of ethanol [65-68].

While there is strong evidence that voluntary exercise can influence consumption of ethanol, the mechanisms responsible for this interaction remain unclear. The mesolimbic dopaminergic (DA) pathway has been implicated in both ethanol consumption and exercise behaviors [83, 129]. Both exercise and ethanol consumption acutely induce DA release in the striatum [83, 133-135]. The mesolimbic DA pathway is composed of DA neurons originating in sub-regions of the midbrain: substantia nigra (SN) and ventral tegmental area (VTA). These neurons project to the striatum—caudate-putamen and nucleus accumbens—as well as to regions of the frontal cortex.

Also important is the hippocampus, which modulates the role of the striatum based on contextual learning. We examine the gene expression of six genes important in regulating this pathway, and previously associated with exercise and/or ethanol consumption. Table 3.1 provides a summary of the expression patterns, functions, reasons for inclusion in the study, and references for these genes.

This study was designed with two aims. First we wanted to replicate the phenomenon of hedonic substitution, and second to investigate mesolimbic DA pathway gene expression plasticity in response to access to ethanol and wheel running that may account for some of the behavioral differences.

3.3 Materials and Methods

3.3.1 Statement on animal care

This study was conducted with approval from the Institutional Animal Care and Use Committee at the University of Colorado, Boulder (Boulder, Colorado) following guidelines established by the Office of Laboratory Animal Welfare. All possible measures were taken to minimize animal discomfort.

3.3.2 Animals

Animals were bred and housed at the Specific Pathogen Free facility, operated by the Institute for Behavioral Genetics at the University of Colorado, Boulder (Boulder, Colorado). Female C57BL/6Ibg mice aged 60-90 days were used for these experiments. Animals were individually housed in polycarbonate cages (30.3 x 20.6 x 26 cm) on a 12-hour light/dark cycle with lights on at 7:00 AM. Room temperature was

Table 3.1 List of genes assayed for expression, and relevant details.

Gene name	Translated protein	Brain expression[a]	Function[b]	Reason for inclusion[c]	References
Th	Tyrosine hydroxylase	Midbrain (ventral tegmental area, substantia nigra), and Pons (locus coerulus)	Rate-limiting enzyme in production of dopamine, hydroxylizes tyrosine into L-DOPA	Implicated in exercise and ethanol behaviors	[82, 136-144]
Slc18a2	Vesicular monoamine transporter 2	Midbrain (ventral tegmental area, substantia nigra, raphe nuclei) and Pons (locus coerulus)	Packaging of cytosolic dopamine into synaptic vesicles to facilitate release	SNPs associated with ethanol behavior, and with locomotor behavior. Knockout mice (+/-) drink more.	[43, 145-148]
Slc6a3	Dopamine active transporter	Midbrain (ventral tegmental area and substantia nigra)	Reuptake of dopamine from the synapse	SNPs associated with ethanol behavior	[43, 142, 146, 149-152]
Drd2	Dopamine receptor D2	Midbrain, striatum, and cortex	G-protein coupled receptor - signaling cascade decreases adenylyl cyclase	Implicated in exercise and ethanol behaviors	[39-41, 82, 83, 142, 153-155]
Drd1a	Dopamine receptor D1	Striatum, cortex, olfactory tubercules, olfactory bulbs	G-protein coupled receptor - signaling cascade activates adenylyl cyclase	Implicated in exercise and ethanol behaviors	[82, 142, 156]
Bdnf	Brain-derived neurotrophic factor	Many regions, but highly expressed in hippocampus	Nerve growth factor important for cell survival and proliferation	Increased after exercise, may play role in exercise neuroprotection from binge ethanol	[83, 129, 157-163]

[a]Brain regions where mRNA is expressed
[b]Major function of translated protein
[c]Criteria for inclusion consisted of prior association with exercise or ethanol-related behaviors. Association could be from polymorphisms, changes in mRNA and/or protein expression, or changes in behavior due to pharmacological or genetic manipulation.

maintained between 23 and 24.5°C. All mice had *ad libitum* access to standard chow (Harlan Laboratories, Indianapolis, Indiana) and water. Animals were monitored daily and body weights were recorded every 4 days. Food was weighed every 4 days, on the same schedule as body weights.

3.3.3 Behavioral paradigm

Mice were tested using a previously established paradigm that lead to differences in ethanol consumption when given access to a free running wheel [63]. The four conditions (n=15/condition) included cages with 1) water only, 2) 1 bottle of water and 1 bottle of ethanol (two-bottle choice), 3) water and ethanol two-bottle choice with a running wheel, and 4) water only with a running wheel. The protocol lasted 16 days. Mice housed with a running wheel (diameter 24.2cm, Harvard Apparatus, Holliston, Massachusetts) had 24-hour access to the wheel for all 16 days. Wheel revolutions were counted using a magnet and magnetic switch (Harvard Apparatus) and recorded daily. Mice housed with ethanol two-bottle choice progressed as follows: water only for days 1-3, 3% ethanol (v/v) for days 4-5, 7% ethanol for days 6-7, and 10% ethanol for days 8-16 (Table 3.2). The side of the cage the bottles were on was alternated every two days. Individual consumption of water and ethanol (if applicable) were recorded daily. On day 16 during the second hour of the light cycle, mice were sacrificed by cervical dislocation. Groups of 5 mice were staggered to start the protocol every 2 days so as to minimize the variation in tissue collection times on day 16. Daily measurements of wheel revolutions (1 day each for 4 mice), water (1 day for 1 mouse) and ethanol consumption (1 day each for four mice) are missing due to sporadic equipment failure

Table 3.2 2x2 Behavioral paradigm for wheel running exposure and ethanol consumption.

	Days 1-3	Days 4-5	Days 6-7	Days 8-16
Running	Water only	3% ethanol & water	7% ethanol & water	10% ethanol & water
	Water only			
Sedentary	Water only	3% ethanol & water	7% ethanol & water	10% ethanol & water
	Water only			

(i.e. switches detecting wheel magnets could be bumped out of alignment or fluid tubes could leak if stopper seal was not secured tight enough). These missing values were imputed as the average of the preceding and following days.

3.3.4 Saccharin control group

In addition, 10 mice were housed with two-bottle choice water and saccharin in two cage conditions (n=5/condition), either with or without wheel in cages described above. After water only for days 1-3, a 0.033% saccharin solution was added for days 4-16 [164, 165]. This concentration was sufficient to produce approximately 95% preference in two-bottle choice. The side of the cage the bottles were on was alternated every two days and individual consumption of water and saccharin were recorded daily.

3.3.5 Quantitative real-time polymerase chain reaction

Whole brains were removed and dissected for midbrain, striatum, hippocampus, and cortex and stored in RNALater™ (Ambion, Foster City, California) at -20°C. Total RNA from dissected regions was extracted using EZNA Total RNA Kit II (Omega Bio-tek, Norcross, Georgia). Quality and quantity of RNA were determined by gel electrophoresis and NanoDrop™ spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts). A260/A280 was determined to be excellent in each case (>1.8). Total mRNA was reverse transcribed using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, California). For real-time quantitative PCR, we used Taqman™ primers and probes (Applied Biosystems) for the following genes: *Bdnf* (Mm04230607_s1), *Drd1a* (Mm01353211_m1), *Drd2* (Mm00438545_m1),

Th (Mm00447557_m1), *Slc6a3* (Mm00438388_m1), and *Slc18a2* (Mm00553058_m1). Endogenous genes *Gapdh* (4352339E) and *Actnb* (4352341E) were used for control. Real-time quantitative PCRs were performed using an ABI 7900HT (Applied Biosystems) running Sequence Detection Systems software (SDS v2.3, Applied Biosystems). All target genes were normalized using the $2^{-\Delta\Delta Ct}$ method [166, 167].

3.3.6 *in situ* hybridization

Whole brains were removed and flash frozen in isopentane on dry ice and stored at -70°C. Brains were sectioned coronally into 14 micron slices using a cryostat (Leica, Wetzlar, Germany), thaw mounted on poly-L-lysine coated glass slides (ThermoFisher Scientific) and stored at -70°C. We used previously established method for *in situ* hybridization of radiolabeled antisense riboprobes [168]. Briefly, probes were transcribed *in vitro* with ^{35}S-UTP (PerkinElmer, Waltham, Massachusetts) as the sole source of UTP. Constructs for each gene, cloned into pT3T7 transcription vectors, were acquired from ThermoFisher Scientific: *Drd1a* – EMM1032-613237 (600bp), *Bdnf* – EMM1032-607279 (800bp), Slc18a2 – EMM1032-591860 (1500bp). All vectors were linearized with EcoRI (New England Biolabs, Ipswich, Massachusetts) and transcribed using T3 RNA polymerase (Promega, Fitchburg, Wisconsin). Hybridizations were performed within 1 day of transcription.

On the day of hybridization, after warming to room temperature, tissue was first fixed with a 4% paraformaldehyde solution (15min), rinsed with 1x phosphate buffered saline (3x5min), then rinsed with 0.1M TEA (2min). Next the tissue was acetylated with 0.25% acetic anhydride in 0.1M TEA (15min) and then dehydrated in graded ethanol

solutions of 50%, 70%, 95%, 100% and 100% (3min each). Radiolabeled riboprobes were diluted in a hybridization buffer containing 50% formamide, 10% dextran sulfate, 300mM NaCl, 10mM Tris, 1mM EDTA, and 1x Denhardt's solution, and ~100µL were pipetted onto a 24mm x 60mm coverslip, then placed upside down covering tissue. Coverslips were sealed to slides using DPX mountant (Sigma-Aldrich, St. Louis, Missouri). Tissue sections were hybridized with riboprobe solution for 16 hours at 60°C. After hybridization, tissue section slides were washed with 4x saline sodium citrate (SSC) before being treated with RNase A (20µg/mL) for 1 hour at 37°C. Then tissue sections were desalted by incubation in graded SSC solutions (all with 1mM DTT) to a final stringency of 0.1xSSC at 65°C. Finally, sections were dehydrated with graded ethanol solutions, dried, and exposed to PhosphoScreens (Packard, Meriden, Connecticut) for at least 1 week. Slides for every mouse for each riboprobe were assayed at the same time to allow for direct comparisons between mice.

In order to relate the intensity of each screen image to a relative measure of tissue radioactivity, tissue standards containing known amounts of ^{35}S were exposed along with tissue on each film. Tissue standards were prepared by mixing measured amounts of isotope with a homogenate prepared from whole brain. Actual concentrations of radioactivity were measured in weighted aliquots. The ^{35}S standards contained from 0 to 25 nCi/mg. Ten standards were used for each isotope, and were used to construct standard curves relating optical density and a measure of radioactivity (counts per minute per mg).

Exposed PhosphoScreens were imaged with a Cyclone PhosphoImage reader (Packard), and image .tif files (600 dpi) were imported into the OptiQuant analysis suite

(Packard). Slides were de-identified and brain regions of interest were circled as well as background. At least 3, and as many as 20 measurements were taken from each animal, and the values obtained were averaged for each animal.

3.3.7 Statistical analyses

A one-way repeated measures ANOVA was used to identify group differences in ethanol consumption (runners vs. non-runners). A one-way repeated measures ANOVA was used to identify group differences in daily wheel revolutions (drinkers vs. non-drinkers). A two-way repeated measures ANOVA was used to determine group differences in body weight (2x2 drinkers vs. non-drinkers and runners vs. non-runners). For repeated measures ANOVAs, missing daily values were imputed from the average of the previous and following days' values. A two-way ANOVA was used to determine group differences due to cage and fluid for food consumption data and for gene expression data (2x2, drinkers vs. non-drinkers and runners vs. non-runners). Average ΔC_t values for each mouse were used for RT-PCR. Average CPM/mg values for each mouse were used for *in situ* hybridization data. Repeated measures ANOVAs were calculated using SPSS v20, two-way ANOVAs were calculated using R v2.15.2 (www.r-project.org).

3.4 Results

3.4.1 Mice

Body weights increased over the course of 16 days (Figure 3.1, $F_{3,168}$=29.32, $p<0.001$) but there were no main effects of ethanol or running. There was a slight

difference in the amount of food consumed, with significant main effects observed for both ethanol and a running wheel (Figure 3.2). Mice that had access to ethanol consumed less food than mice that only had access to water (3.52 ± 0.09 g/day vs. 3.85 ± 0.08 g/day, respectively; $F_{1,56}$=8.393, $p<0.01$). Mice with access to a running wheel consumed more food than mice housed in empty cage (3.83 ± 0.09 g/day vs. 3.54 ± 0.08 g/day, respectively, $F_{1,56}$=6.811, $p<0.05$).

3.4.2 Voluntary running and ethanol consumption

As expected, mice ran a considerable distance each day, averaging 7144 revolutions per day, equaling 5431 meters per day. There was a slight increase in daily revolutions over the course of 16 days (Figure 3.3, $F_{15,420}$=2.8, $p<0.001$) There was no significant difference in number of revolutions between mice with access to water only and mice with access to ethanol.

Mice with access to running wheel significantly consumed (g/kg; $F_{1,28}$=11.6, $p<0.01$) and preferred ($F_{1,28}$=30.7, $p<0.001$) less ethanol than mice housed in an empty cage over the course of 16 days (Figure 3.4, a and b).

3.4.4 Saccharin control group

In the ten mice in the saccharin control experiment there was no significant change in body weight over the 16 days, nor was there any effect of access to a running wheel (Figure 3.5) . Access to a running wheel did not significantly change saccharin

Figure 3.1 Body weight over the time course of the experiment.

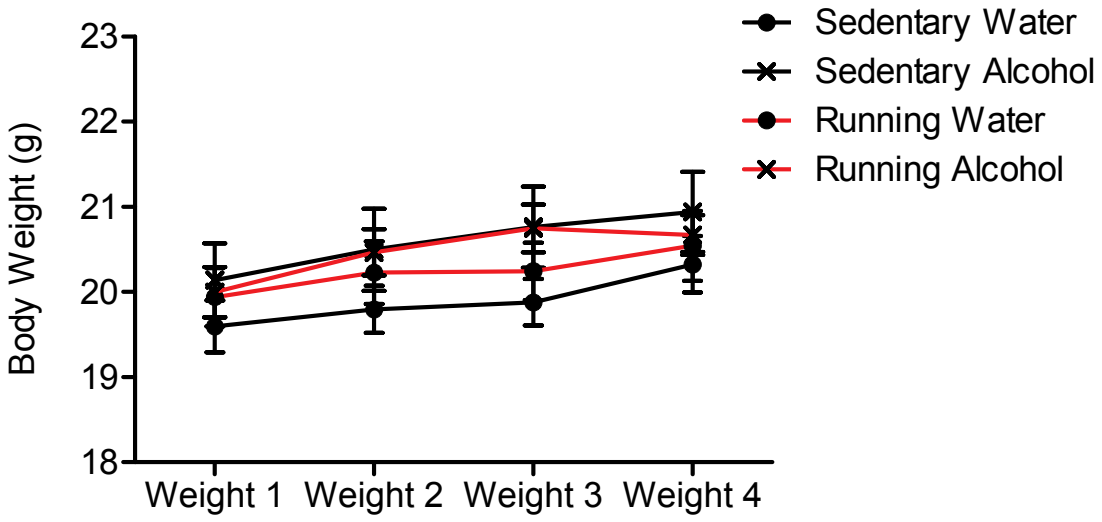

Figure 3.1 shows the average body weight (g) of each group of mice, the two running groups are red lines, the two ethanol consuming groups are the X's. Body weight for all groups increased over the course of the experiment ($F_{3,168}$=29.32, $p<0.001$), and there were no significant differences between groups. Means ± SEM are reported.

Figure 3.2 Average daily food consumption.

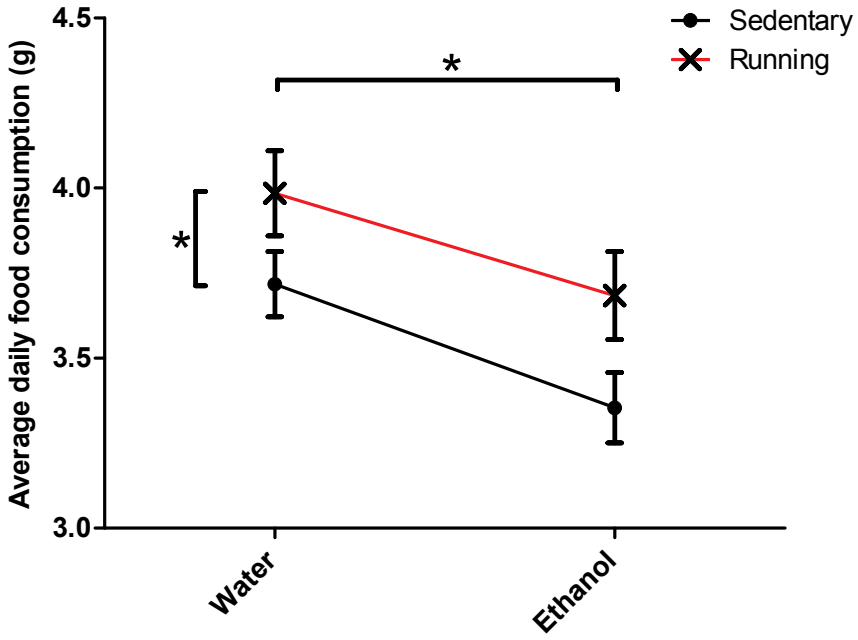

Figure 3.2 shows the average daily food consumption (g) for each group of mice. The red line indicates the mice were running, the black line indicates the mice were sedentary. Two-way ANOVA shows that running mice consumed more food ($F_{1,56}$=6.811, $p<0.05$) and that ethanol consuming mice ate less food ($F_{1,56}$=8.393, $p<0.01$). Means ± SEM are reported.

Figure 3.3 Average daily wheel revolutions.

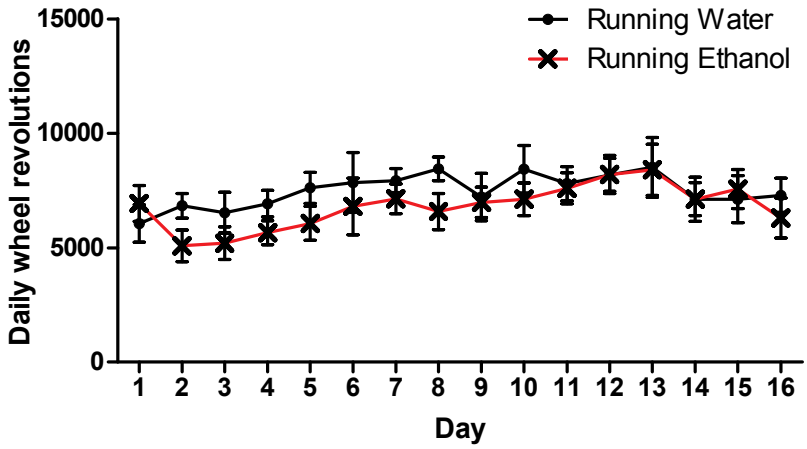

Figure 3.3 shows average daily running wheel revolutions for mice with water only and mice with two-bottle choice ethanol over 16 days. Average number of revolutions for water only mice (black line, n=15, 7486 ± 590 revolutions/day) and for ethanol-drinking mice (red line, n=15, 6798 ± 584 revolutions/day). Mean ± SEM are reported.

Figure 3.4 Average daily ethanol consumption.

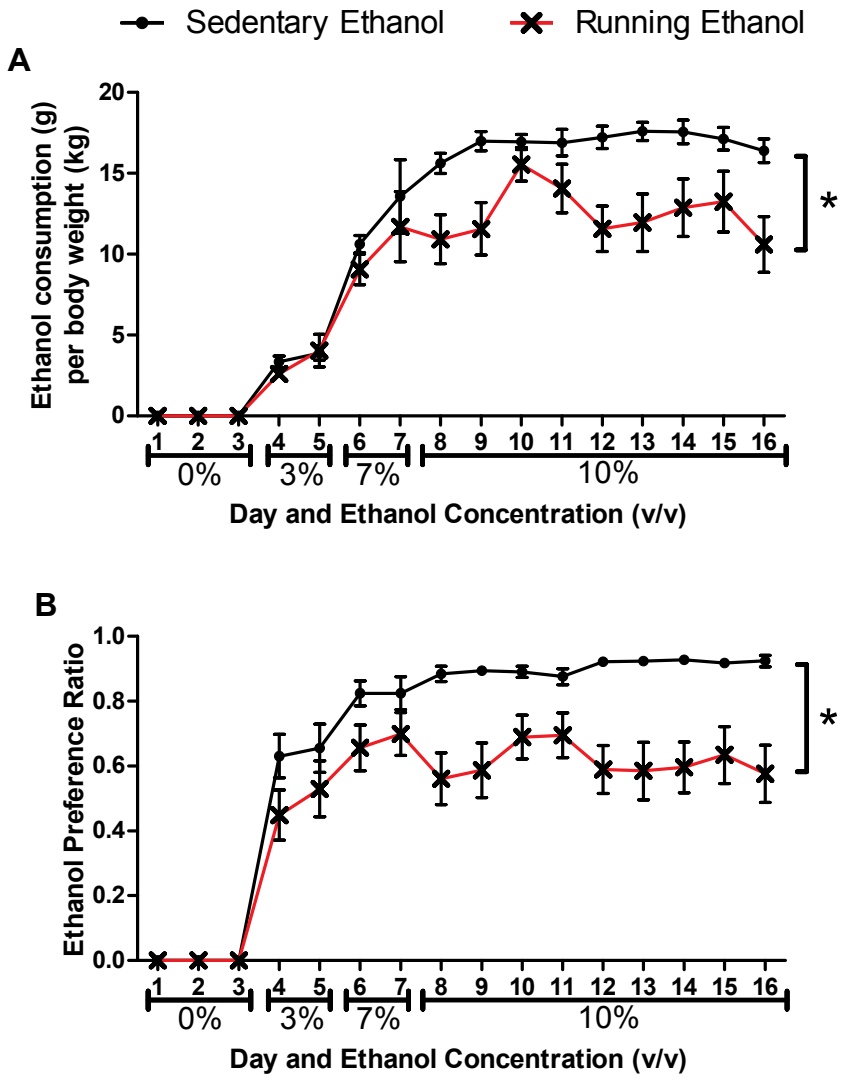

Figure 3.4 shows average daily ethanol consumption for sedentary mice (black line) and mice with access to a running wheel (red line) over 16 days. Ethanol consumption is shown as average amount of ethanol consumed per body weight (A) and as an ethanol preference ratio (B) defined as volume of ethanol fluid consumed divided by total fluid consumed. Ethanol concentrations (v/v) for each day are reported on the x-axis. Mean ± SEM are reported.

Figure 3.5 Average daily saccharin consumption.

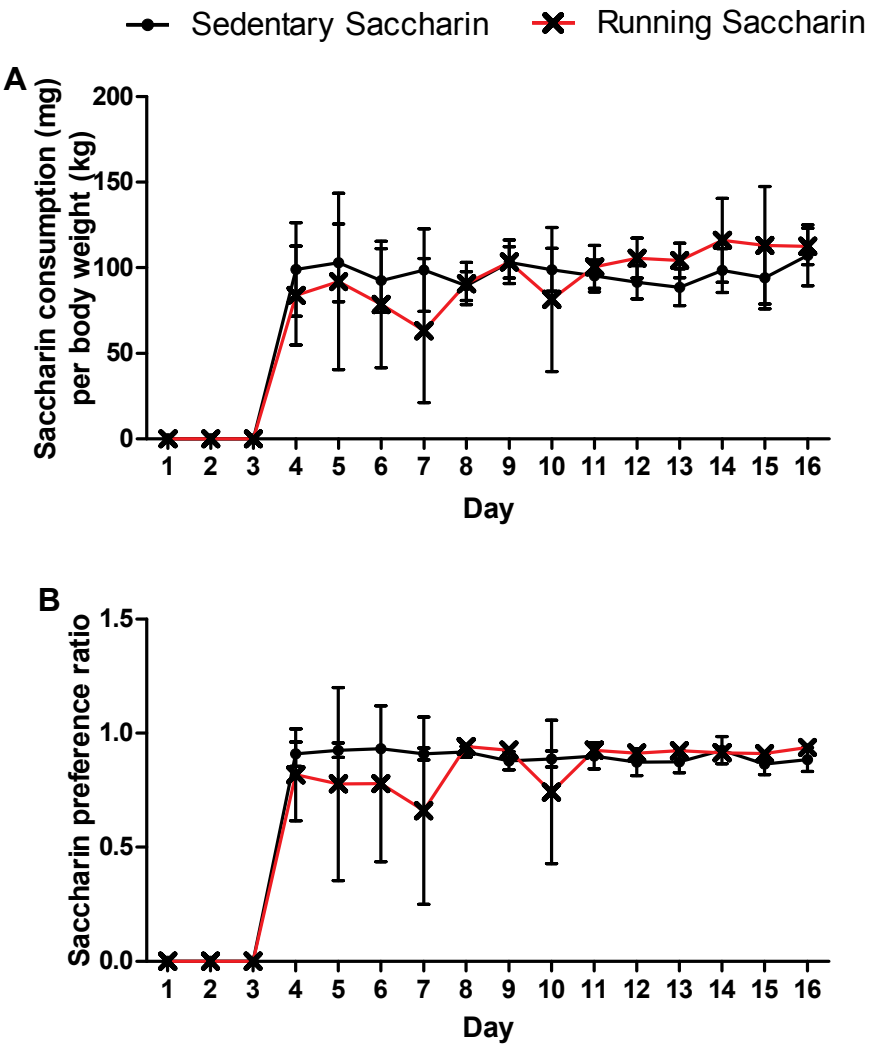

Figure 3.5 shows average daily saccharin consumption for sedentary mice (black line) and mice with access to a running wheel (red line) over 16 days. Saccharin consumption is shown as average amount of saccharin consumed per body weight (A) and as a saccharin preference ratio (B) defined as volume of saccharin fluid consumed divided by total fluid consumed. A 0.033% (v/v) saccharin solution was used for each day 4-16. Mean ± SEM are reported.

consumption as measured by milligrams saccharin per kilogram body weight or as a saccharin preference ratio.

3.4.5 Quantitative real-time polymerase chain reaction

There were no significant differences in gene expression between groups for *Th, Drd2,* and *Slc6a3* (Figure 3.6, a-e). In addition, there were no differences in gene expression for *Drd1a* when measured in the cortex (Figure 3.6h). In the midbrain, there was a main effect of ethanol availability on expression of *Slc18a2* (Figure 3.6f, $F_{1,16}$=18.9, p<0.001), ethanol-consuming mice showing increased expression compared to the group that only had access to water. In the striatum, there was a main effect of ethanol availability on the expression of *Drd1a* (Figure 3.6g, $F_{1,16}$=6.9, p<0.05), with ethanol-consuming mice showing decreased expression levels. In the hippocampus, there were significant main effects of running wheel availability ($F_{1,17}$=6.3, p<0.05) and ethanol availability ($F_{1,17}$=5.5, p<0.05) on *Bdnf* expression (Figure 3.6i). Running mice had increased expression of *Bdnf*, while ethanol-drinking mice had decreased expression.

3.4.6 *in situ* hybridization

There were different expression patterns for *Slc18a2* and *Drd1a* when measured using *in situ* hybridization. Contrary to expression levels detected through qRT-PCR, there was an ethanol x running wheel interaction effect on the expression of *Slc18a2* in both midbrain sub-regions: the substantia nigra ($F_{1,18}$=5.2, p<0.05) and the ventral tegmental area ($F_{1,18}$=5.6, p<0.05, Figures 3.7, a and b). In empty cages, ethanol-

Figure 3.6 Relative gene expression as measured by qRT-PCR.

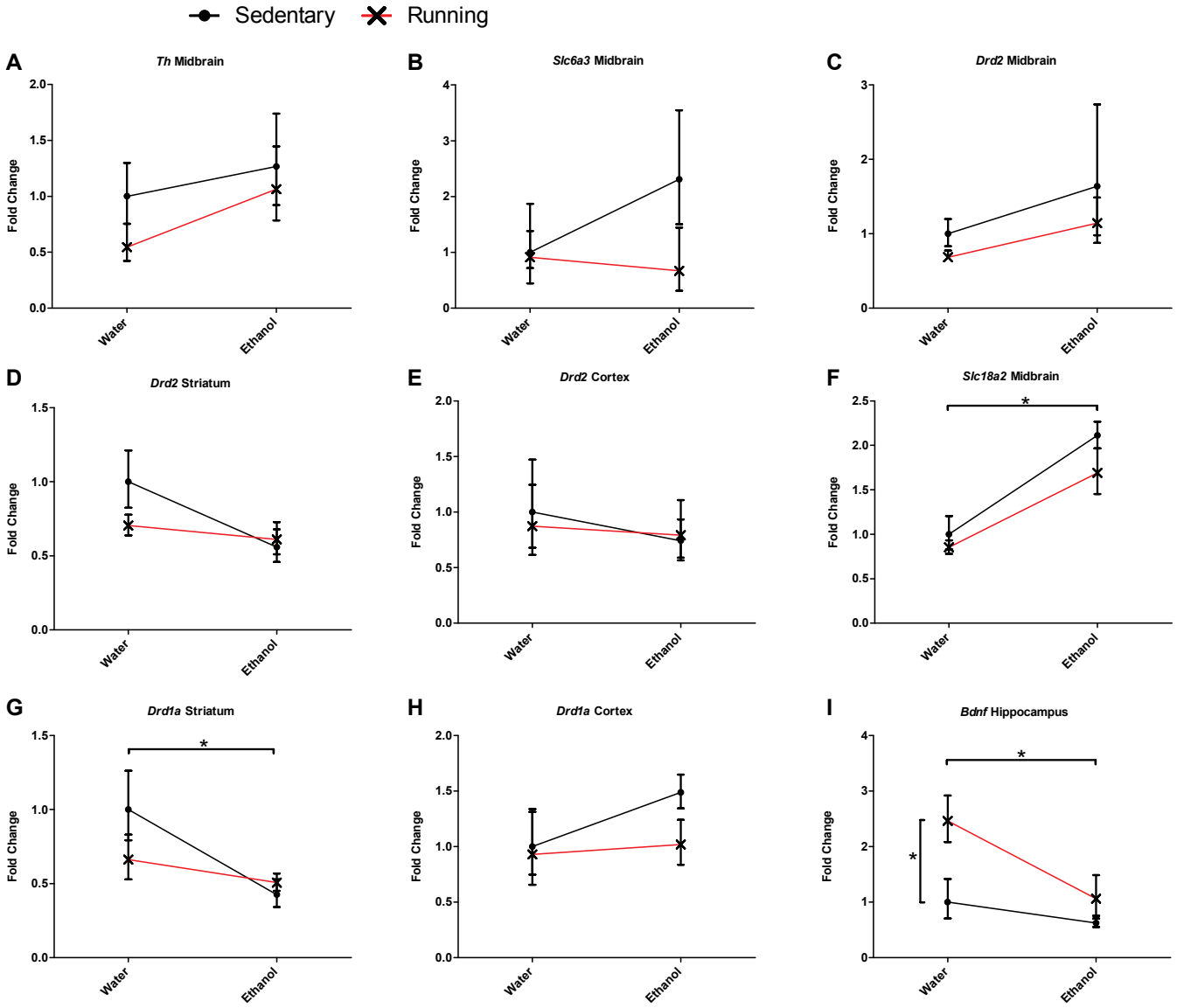

Figure 3.6 shows relative mRNA expression levels as measured by quantitative real-time PCR for (A) *Th* (midbrain), (B) *Slc6a3* (midbrain), (C-E) *Drd2* (midbrain, striatum, and cortex), (F) *Slc18a2* (midbrain), (G-H) , *Drd1a* (striatum and

cortex), and (I) *Bdnf* (hippocampus). Main effects due to availability of ethanol were observed in midbrain *Slc18a2*, striatal *Drd1a*, and hippocampal *Bdnf*. Main effects due to availability of a running wheel were observed in hippocampal *Bdnf*. There were no significant interaction effects. Expression levels are shown as mean fold change ± SEM relative to sedentary/water only group for each gene.

Figure 3.7 Relative gene expression as measured by *in situ* hybridization.

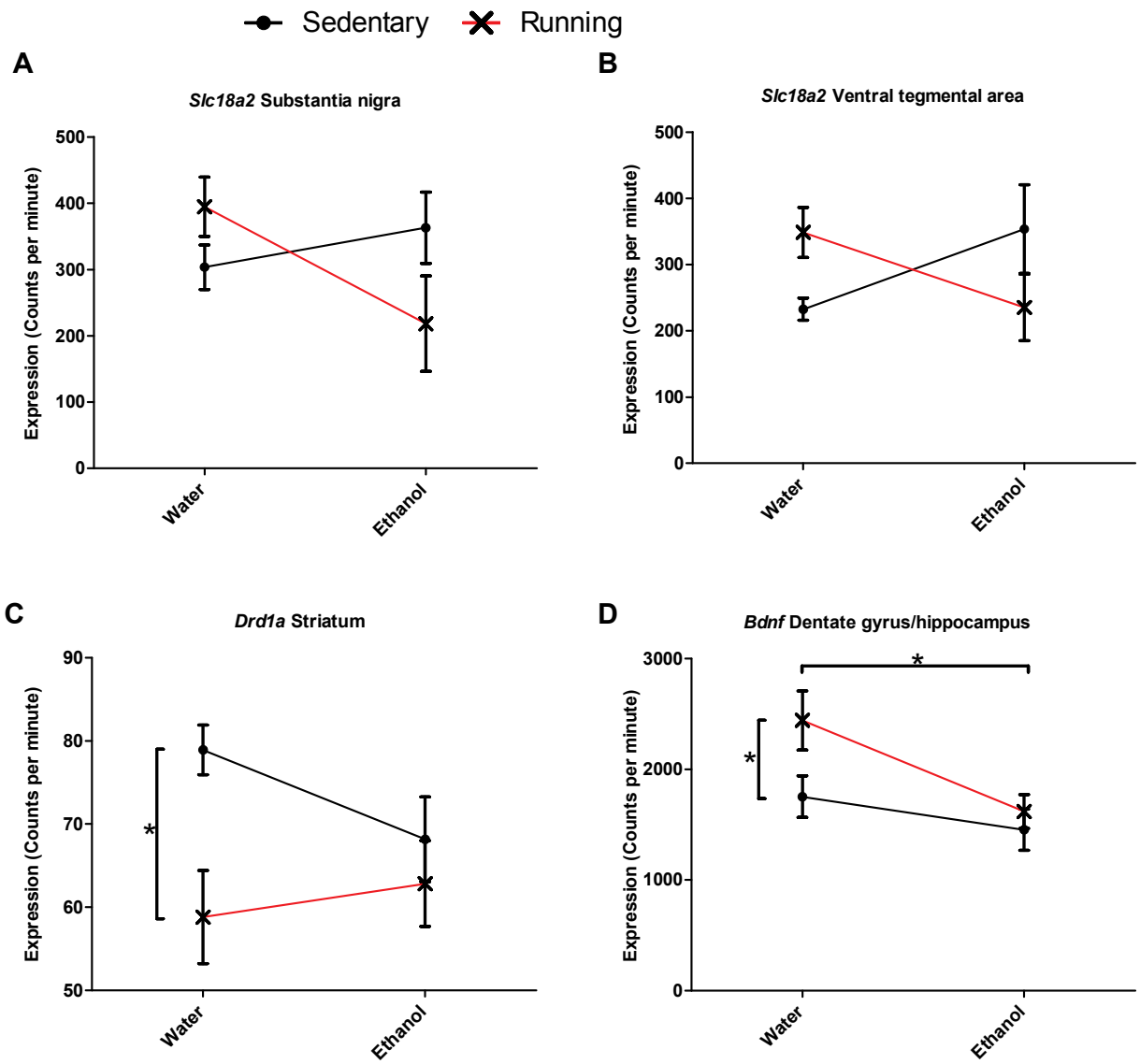

Figure 3.7 shows results of *in situ* hybridization showing mRNA expression levels for *Slc18a2* in midbrain subregions: (A) substantia nigra (SN) and (B) ventral tegmental area (VTA), *Drd1a* in (C) striatum (ST), and (D) *Bdnf* in hippocampus

(HC). Significant interaction effects were observed in both midbrain regions for *Slc18a2.* Significant main effects due to availability of running wheel were observed in striatal *Drd1a* and in hippocampal *Bdnf.* A main effect due to availability of ethanol was observed in hippocampal *Bdnf.* Values shown (mean ± SEM) have been converted to counts per minute.

drinking mice had increased expression of *Slc18a2* compared to water-only mice, while in cages with running wheels; ethanol-drinking mice had decreased expression of *Slc18a2*. In the striatum, there was a main effect of running wheel availability on expression of *Drd1a* (Figure 3.7c). Mice with access to running wheels showed decreased expression *Drd1a* compared to mice without access ($F_{1,20}$=7.0, $p<0.05$). There was no effect of availability of ethanol on *Drd1a* expression. In concordance with expression levels detected through RT-PCR, there were significant main effects of access to running wheel ($F_{1,20}$=4.4, $p<0.05$) and access to ethanol ($F_{1,20}$=7.7, $p<0.05$) on expression of *Bdnf* in the dentate gyrus of the hippocampus (Figure 3.7d). Access to a running wheel increased expression, while access to ethanol, decreased expression. There was no significant interaction effect. Representative images from each *in situ* hybridization assay are shown in Figure 3.8.

3.5 Discussion

3.5.1 Evidence for hedonic substitution

This study provides additional evidence for the behavioral interaction between voluntary exercise and ethanol consumption. Using a similar protocol as in Ehringer et al (2009), we observed decreased ethanol consumption and preference in female C57Bl/6Ibg mice than in their non-running counterparts [63]. Ehringer (2009) demonstrated that the substitution effect was due specifically to a running wheel, and not to a locked wheel, at least in females. They went on to show that ethanol metabolism was unchanged due to running. The current study shows running on an exercise wheel failed to reduce consumption and preference for a saccharin solution,

Figure 3.8 Representative coronal sections showing areas of gene expression for *in situ* hybridations.

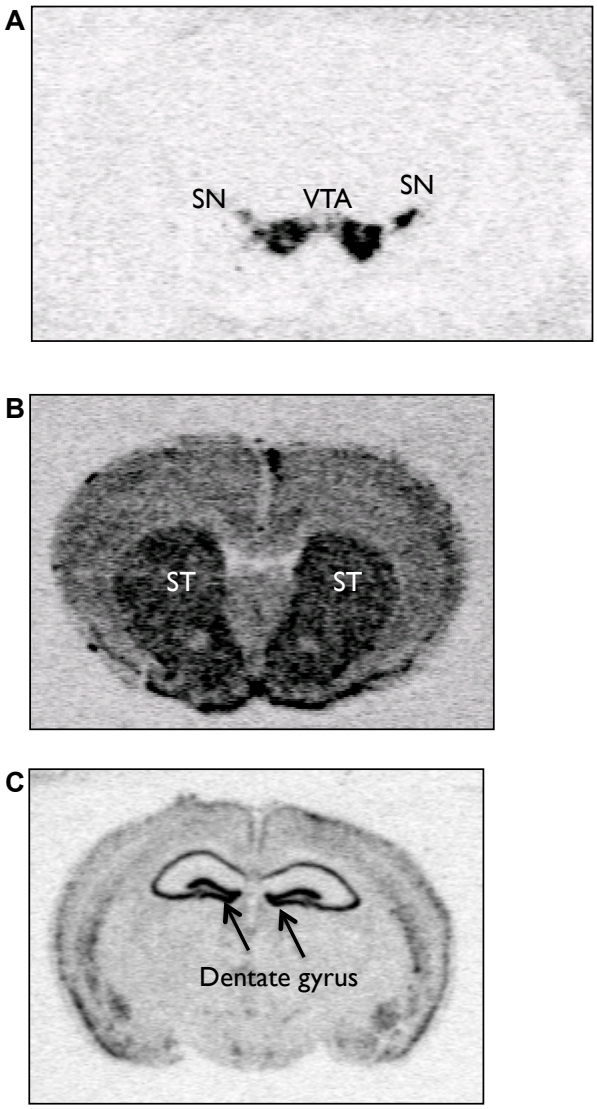

Figure 3.8 shows representative images of each *in situ* hybridization assay. For *Slc18a2* (A), both the substantia nigra (SN) and ventral tegmental area (VTA) were quantified. For *Drd1a* (B), the whole striatum (ST) was quantified. For *Bdnf* (C), the dentate gyrus of the hippocampus was quantified.

suggesting that the effects of exercise may not be sufficient to alter all rewarding behaviors.

3.5.2 Transcriptional changes in mesolimbic reward pathway

In addition to providing further evidence for the behavioral interaction, this study attempted to elucidate the underlying molecular basis for hedonic substitution, utilizing measures of gene expression. Although exercise has been and is currently being used as a behavioral intervention for alcohol use disorders[65-68], a more thorough understanding of the mechanisms for its usefulness could provide a framework for more effective interventions. The implication of the mesolimbic DA pathway in the regulation of voluntary exercise and ethanol consumption provided a foundation for hypothesizing which genes could play a role. We selected genes based on their role modulating this pathway, and based on prior implications in exercise and ethanol behaviors.

We examined the expression of six genes, all actors in the mesolimbic DA pathway, first by quantitative real-time PCR, then if a main effect was observed, by *in situ* hybridization. In the midbrain we were unable to detect changes in expression of *Th, Drd2,* and *Slc6a3.* Although several groups have shown an increase in *Th* in response to exercise and ethanol, protocol and organism differences may explain the discrepancy in findings. Greenwood (2011) saw differences in the caudal third of the VTA of Fischer 344 rats after 6 weeks of running, with no differences in the mid and rostral portion, and no differences reported for SN [82]. Other *in vitro* studies observed increases in *Th* in response to ethanol administration [169]. Polymorphisms in *Slc6a3* have been associated with alcohol use disorders in humans [150]; however, knockout

mice have similar ethanol drinking behaviors, at least in females, as their wild-type controls [43]. Therefore, it is unclear whether the expression of *Slc6a3*, affects voluntary ethanol or exercise behavior. Polymorphisms in *Drd2,* as well as changes in expression (human and animal), have been implicated in both behaviors [39, 82, 83, 142, 153-155]. Non-significant differences between ethanol and/or exercise groups in our study may reflect differences in length of protocol or animal model.

In the midbrain, mRNA expression measured using qRT-PCR assays of *Slc18a2* appeared to increase in response to consumption of ethanol. However, when looking at more fine-grained patterns in the VTA or SN using the *in situ* technique, the main effect of ethanol was abolished, and we observed an interaction effect. In the absence of a running wheel, ethanol led to increased expression, while in running mice, ethanol led to decreased expression. Other midbrain regions expressing *Slc18a2,* such as the raphe nuclei in the caudal midbrain, may account for some of these differences, and highlight the importance of more targeted regional selections. The interaction effect observed in the VTA and SN is particularly interesting, and we speculate that in the absence of running, there could be increased DA release due to ethanol, which facilitates the need for higher expression of *Slc18a2*. This could be tested in future studies.

In the striatum, expression of *Drd1a* showed different patterns of response dependent of the method of detection. In the qRT-PCR experiments, *Drd1a* showed decreased expression in response to ethanol, while this decrease was observed in the wheel running condition for *in situ* assays. The mice in this study were voluntarily drinking ethanol, and therefore not ethanol-dependent, however Contet et al (2011) observed increased *Drd1a* expression in ethanol-dependent mice [156]. Our

observation of reduced expression measured by ISH is consistent with work by Knab et al (2009), who observed lower baseline *Drd1a* expression in a high-running strain of mice compared to a low-running strain [142].

The hippocampus provides contextual information to the striatum based on prior associated experiences [129]. In one of the few studies examining the influence of exercise on ethanol behaviors, Leasure and Nixon (2009) demonstrated exercise's ability to protect hippocampal cells from the effects of binge ethanol consumption [159]. These findings complement other work showing the ability of exercise to initiate neurogenesis in the hippocampus [158, 163]. Our data, consistent using both qRT-PCR and ISH, show that 16 days of voluntary exercise increases hippocampal *Bdnf* expression and 16 days of voluntary ethanol consumption decreases expression, which suggests changes in neuronal structure and neurogenesis. Alterations in the hippocampus due to exercise may influence the signaling between the striatum and hippocampus, affecting the reward response to ethanol drinking.

3.5.3 Conclusions

These data reaffirm the hedonic substitution narrative, and make the first attempt at identifying the underlying genetic changes that occur to influence this interaction. Of the genes where differences were observed, *Slc18a2* expression in the VTA and SN responds differently to ethanol depending on the presence or absence of a running wheel, and *Bdnf* expression in the hippocampus changes in response to both running and ethanol. *Drd1a* in the striatum may also be responsive to both running and ethanol, although most likely just running. This suggests that multiple genes and brain regions

are important in regulating hedonic substitution, and supports the idea that the mesolimbic DA pathway plays an important role. Future studies should focus on global gene expression to identify other genes as well as assessing whether observed mRNA changes correspond to similar protein changes. It will be useful also to expand this behavioral model to include testing whether running produces similar effects on voluntary consumption of other drugs of abuse.

Chapter 4

Identification of candidate genes and pathways involved in the hedonic substitution of exercise for ethanol consumption

Todd M Darlington and Marissa A Ehringer

In preparation.

4.1 Abstract

The behavioral interaction known as hedonic substitution has been observed and replicated, and again we show mice decrease ethanol consumption and preference when given access to a running wheel. However, the despite attempts at identifying the underlying neurobiological mechanism, this remains unknown. To identify candidate genes and pathways involved in hedonic substitution, we quantitatively sequenced mRNA from the striatum of female C57BL/6J mice. There were four groups of mice, control, access to two-bottle choice ethanol, access to a running wheel, and access to both two-bottle choice ethanol and a running wheel. We identified many differentially expressed genes, including several in ethanol preference quantitative trait loci that are differentially expressed in response to running. Furthermore, we conducted Weighted Gene Co-expression Network Analysis and identified putative exercise responsive gene networks, with one network implicating a role for glial cells. We identify roles for potassium channel genes as well as other candidate genes, *Ttr, Stx1b*, and *Oprm1* in regulating hedonic substitution. Because many of the genes and functional groups have been previously identified in studies of initial sensitivity to ethanol, we propose that exercise may induce a change in sensitivity, which affects ethanol consumption. Furthermore, these results provide a rich resource for studies involving transcriptional changes in gene networks in response to ethanol consumption and wheel running.

4.2 Introduction

Although the prevalence of alcohol use disorders remains high, there are relatively few treatment options available [1-4, 6]. The concept of hedonic substitution, the replacement of one rewarding behavior with another, is a promising area of research. Exercise has been used in the past to help reduce ethanol intake in heavy drinkers [65-68], and has consistently interacted with ethanol consumption in laboratory animal studies [58-63], but little is known about the neurobiology of this interaction. Chapter 3 was the first study attempting to identify transcriptional changes underlying hedonic substitution, and the candidate genes *Slc18a2* in the midbrain and *Bdnf* in the hippocampus were both found to respond differently to ethanol consumption and voluntary exercise. The striatum plays an important role in the mesolimbic dopaminergic pathway, processing and integrating input from a number of other brain regions. Therefore, it seems probable that multiple transcriptional events occur in the striatum that may provide insight into hedonic substitution.

The mesolimbic dopaminergic pathway consists of dopaminergic neurons in the ventral tegmental area (VTA) of the midbrain that project to the nucleus accumbens (NAc) in the ventral striatum, releasing dopamine (DA) upon stimulation. Both ingestion of ethanol and voluntary exercise facilitate increased DA levels in the NAc [83, 133, 170]. However, there is increasing evidence that in the striatum, the whole striatum is involved in developing addiction [129, 171]. While initial exposure to hedonic stimuli stimulates the shell of the NAc and feeds back to the VTA, interactions between the shell and the core of the NAc induce conditioned reinforcement to the stimuli. Furthermore, animals will respond to direct stimulation of substantia nigra as well as

VTA, suggesting an acute nigrostriatal role in hedonia [172-174]. In heavy drinking humans, there is greater activation in the dorsal striatum than in the ventral striatum when presented with drinking-related cues [175]. These studies demonstrate the importance of inclusion of the whole striatum when considering ethanol related changes.

This study is designed to identify candidate genes for hedonic substitution by examining the striatal transcriptome. Priority will be given to genes located in ethanol preference quantitative trait loci (QTL) on chromosomes 2 and 9 [31-37]. The current study utilizes RNA-Sequencing to compare the transcriptional responses to voluntary ethanol consumption and wheel running. Furthermore, Weighted Gene Co-expression Network Analysis (WGCNA), an agnostic network analysis tool will be used to identify biologically relevant co-expression networks. RNA-Sequencing is relatively new technology, but using expression data produced from RNA-Seq with WGCNA has been shown to improve network characteristics relative to microarray expression data [88] and has been used successfully with WGCNA to identify biologically relevant co-expression networks related to ethanol sensitivity [85].

4.3 Materials and Methods

4.3.1 Statement on animal care

This study was conducted with approval from the Institutional Animal Care and Use Committee at the University of Colorado, Boulder (Boulder, Colorado) following guidelines established by the Office of Laboratory Animal Welfare. All possible measures were taken to minimize animal discomfort.

4.3.2 Animals

Adult female C57BL/6J mice, bred and housed at the Specific Pathogen Free facility at the Institute for Behavioral Genetics (University of Colorado, Boulder), were used for this study. Mice were group-housed in their home cages on the testing floor for at least 6 days prior to individual housing. On the first day of testing, mice were individually housed in polycarbonate cages with dimensions 30.3cm x 20.6cm x 26cm with cedar wood chips and a bedding square. The room was on a 12 hour light/dark cycle with lights on at 7:00AM. Room temperature and humidity were monitored every day, with temperatures ranging from 23 – 24.5°C and humidity ranging from 20 – 40%. Mice had *ad libitum* access to both water and standard chow (Harlan Laboratories, Indianapolis, Indiana), and were monitored daily. Body weight and food consumption were measured every four days.

4.3.3 Behavioral paradigm

Mice were tested using the protocol described in Chapter 3, and consistent with methods previously described as producing an hedonic substitution effect [63]. Groups of 5 mice were started at a single time, staggered every 2 days. Conditions were randomized between staggered groups. Briefly, mice were house under one of four cage conditions (Table 3.1, page 56, n=6/condition), including cages with: 1) water only and no wheel, 2) water and ethanol two-bottle choice and no wheel, 3) water only with a running wheel, and 4) water and ethanol two-bottle choice with a running wheel. Mice housed with running wheels had continuous 24-hour access to the running wheels

(diameter 24.2cm, Harvard Apparatus, Holliston, Massachusetts) each day of the 16-day protocol. Running wheel revolutions were measured daily using a magnetic switch (Harvard Apparatus) triggered by a magnet on the wheel. Mice with access to two-bottle choice ethanol and water progressed as follows: water only for days 1-3, 3% ethanol (v/v) for days 4-5, 7% ethanol for days 6-7, and 10% ethanol for days 8-16. To prevent a side preference in the drinking bottles the side of the cage the bottles were on was alternated every two days. Volumes of water and ethanol (if applicable) consumed were measured daily. Bottle leakages were determined using a daily outlier test, with a threshold of 2 standard deviations, however none were detected in this study. One daily wheel revolution count was missing from two mice, due to accidental misalignment of magnetic switch. Those two missing values have been imputed from the average of the two-nearest daily values.

4.3.4 RNA extraction and preparation

Immediately after cervical dislocation, brains were removed and the whole striatum was dissected out and placed in RNALater (Ambion, Foster City, California). Total RNA was extracted and purified using Qiagen RNeasy Midi kits (Qiagen, Valencia, California). Quantity and quality were determined using a NanoDrop™ spectrophotometer (Thermo Fisher Scientific, Wilmington, Delaware) and Agilent 2100 BioAnalyzer™ (Agilent Technologies, Santa Clara, California). Ratios of absorbance (260nm:280nm) were shown to be excellent (>1.8). RNA Integrity scores were also shown to be excellent (>8.0). For each sample, five µg of total RNA was used, first for ribosomal RNA (rRNA) depletion using Ribo-Zero™ Magnetic kits (Epicentre

Biotechnologies, Madison, Wisconsin), then poly-A enrichment using Dynabeads® oligo-dT magnetic beads (Invitrogen, Carlsbad, California), both according to kit specifications.

The preparation of the cDNA libraries was performed using ScriptSeq™ V2 RNA-Seq Library Prep kit (Epicentre Biosystems), which generated strand-specific pair-end libraries for quantitative RNA-Sequencing on Illumina platforms. The protocol followed is described in detail at www.epibio.com. Briefly, 50ng of mRNA was fragmented, and then reverse transcribed to single stranded cDNA. This first strand of cDNA was di-tagged with a 58 nucleotide oligomer, before purification with Agencourt AMPure XP beads (Beckman Coulter, Brea, California). Purified, di-tagged single stranded cDNA was then amplified with 15 cycles by polymerase chain reaction using ScriptSeq™ V2 Index Primers (Epicentre Biosystems), designed to add a 6 nucleotide unique barcode to each cDNA in each sample. After amplification, cDNA was purified using Agencourt AMPure XP beads (Beckman Coulter) and shipped to the University of Colorado, Denver, Sequencing Core Facility. Upon arrival, samples were tested for quality and quantity using Agilent 2100 BioAnalyzer™ (Agilent Technologies) and Qubit® 2.0 Fluorometer (Invitrogen).

4.3.5 RNA-Sequencing

Four samples per lane (one per condition, ScriptSeq™ barcodes 4-7) were run on six lanes of an Illumina HiSeq 2000 (Illumina, San Diego, California), pair-end sequenced to 100 nucleotides. After sequencing, the core facility provides de-barcoded reads. Fastq files were assessed for quality using FastQC (v0.3, Babraham Institute).

Using the FASTQ Trimmer (v1.0.0) [176] six nucleotides from the 5' end of each read were trimmed due to base composition bias at those positions. Trimmed reads were aligned to the mouse reference genome (mm9, Ensembl) using TopHat (v2.0) [95], allowing for 2 mismatched bases, up to 10 alignments per read, no mismatches in secondary segment alignment, and only aligning across known exon junctions. To assemble transcripts and generate read counts per transcript, output from TopHat and the annotated reference genome (mm9, Ensembl) was analyzed using Cufflinks (v2.0.2) to construct the minimum number of transcripts that explain the maximum number of reads [98]. Since the sequenced sample had been rRNA depleted and enriched for poly-A mRNA transcripts, a mask file was used to discriminate against alignments in rRNA, tRNA, and small RNA genes. Read counts per transcript was then output to EdgeR (v3.0.8) [177-180], which was used to test for differential expression. For genes to be included in differential expression testing, there had to be at least one aligned read in each sample for that gene. This minimal threshold is consistent with other studies utilizing RNA-Sequencing/EdgeR [88, 126]. EdgeR runs in R/Bioconductor [181, 182] and provides for the statistical analysis of raw count data. EdgeR allows for fitting a general linear model, allowing the inclusion of an interaction term, which is important for interpreting this 2x2 experimental design, and relies on the negative binomial distribution to infer differential expression. P-values are corrected using a Benjamini-Hochberg false discovery rate of 5% (FDR<0.05) [104].

4.3.6 Weighted Gene Co-expression Network Analysis (WGCNA)

The Weighted Gene Co-expression Network Analysis (WGCNA, v1.25.2) provides an agnostic analysis of patterns of gene expression, regardless of treatment condition [101, 103]. Similarly co-expressed genes are clustered into modules, which can then be related to treatment and biological relevance. EdgeR provides normalized read counts for each gene, which were used for a single WGCNA with all 24 samples, similar to the approach used in Chapter 2 [85]. First, a signed similarity matrix was constructed with Equation 2.1. This was converted to a weighted adjacency matrix by a power function (Equation 2.2), determined by a scale-free topology model ($\beta=8$). Therefore, the adjacency matrix contained values from 0 to 1 for each gene, with 0, 0.5, and 1; signifying negative correlation (0-0.5), no correlation (0.5), and positive correlation (0.5-1). The weighted adjacency matrix was converted to topological overlap matrix (TOM, Equation 2.3), then a measure of dissimilarity was generated by 1 - TOM. Genes were clustered based on hierarchical clustering of TOM-based dissimilarity, with the dynamic tree cutting algorithm cutreeDynamic, and the deepSplit option set to 2. Gene clusters with a minimum of 30 genes were identified using a dynamic tree-cutting algorithm, which identified 91 gene clusters (modules). Similar gene modules were merged using the mergeCloseModules command, with a dissimilarity threshold of 0.2 (Pearson correlation greater than 0.8). Merging similar modules resulted in 29 remaining modules used in downstream analysis. Hub genes in each module were determined by ranking each gene by its module membership, calculated by WGCNA. Module robustness was tested in three ways. First, average module adjacencies were calculated and compared to the average adjacencies of randomly sampled “modules” of the same size. One thousand permutations of randomly sampled modules were

generated. Modules were considered robust if average module adjacencies were significantly higher than the randomly generated modules. Second, we repeated this permutation test using average topological overlap, similar to Iancu et al (2012). [88]. Third, the intramodular and extramodular connectivity of each module was calculated and scaled according to module size. Modules with higher scaled intramodular connectivity were considered robust.

To identify experimentally relevant co-expression modules, we took the first principle component of the expression data of each module using the moduleEigengenes command from the WGCNA R-package. The resulting module eigengenes are representative of the gene expression levels for each module, if the module were reduced to a single gene. A two-way analysis of variance of the resulting module eigengene values was used to identify module eigengenes different due to access to a running wheel, access to ethanol, or an interaction effect. Significant p-values were less than 0.05/29=0.0017.

4.3.7 Functional group over-representation

Each set of differentially expressed genes and each WGCNA module were tested for functional group over-representation with the Web-based gene set analysis toolkit (WebGestalt, http://bioinfo.vanderbilt.edu/webgestalt) [105, 106]. Functional groups based on Gene Ontology (GO) [107], Kyoto Encyclopedia of Genes and Genomes (KEGG) [108, 109], and WikiPathways [110, 111]. Furthermore, using a database of genes differentially expressed by cell type—neurons, astrocytes, and oligodendrocytes [112]—we tested whether these cell-type specific genes were over-

represented in the sets of differentially expressed genes or in each WGCNA module, using a hypergeometric test in R.

4.4 Results

4.4.1 Behavior

Over the course of the 16-day protocol, mouse body weights increased, with no effects of access to ethanol or running ($F_{3,60}$=9.4, p<0.001, repeated measures ANOVA). There is a main effect of both access to a running wheel and access to ethanol on food consumption as measured by two-way ANOVA. Mice with access to a running wheel consumed on average slightly more food ($F_{1,20}$=27, p<0.001) and mice with access to ethanol consumed slightly less food on average ($F_{1,20}$=9.7, p<0.01). There was no effect of access to ethanol on average daily running wheel revolutions. Mice with access to a running wheel consumed ($F_{12,120}$=10.2, p<0.001, repeated measures ANOVA) and preferred ($F_{12,120}$=27.9, p<0.001, repeated measures ANOVA) less ethanol than mice without access to a running wheel.

4.4.2 RNA-Sequencing

We quantitatively sequenced 24 striatum samples from 4 groups of mice on the Illumina HiSeq 2000 platform. We generated paired-end reads, 100 nucleotides long, and after trimming 6 bases from the 5’ end, aligned them to the mouse reference genome, masked to preferentially align to protein coding genes. General results from the sequencing and alignment are shown in Table 4.1. 14207 genes were expressed in our samples, meeting the minimum threshold and tested for differential expression.

Table 4.1 RNA-Sequencing and alignment details

Group[a]	Total reads[b]	% mapping[c]	Total read pairs[d]	% mapping[e]
Sedentary Water	50097151 ± 6638016	58.2	49812854 ± 6585960	57.0
Running Water	43570374 ± 7291240	56.1	43338688 ± 7243010	54.6
Sedentary Ethanol	49009532 ± 5654175	54.1	48752299 ± 5623542	53.0
Running Ethanol	54308170 ± 8274994	57.3	54003523 ± 8232762	56.2

[a]Treatment group
[b]Total number of reads, average per treatment group ± SEM
[c]Percentage of total reads aligning at least once, and up to 10 times
[d]Total number of read pairs, average per treatment group ± SEM
[e]Percentage of total read pairs aligning at least once, and up to 10 times

At an FDR<0.05, there were 247 differentially expressed genes (DEGs) due to the access to running wheel (203 with no interaction effect, only a main effect), 37 genes were differentially expressed due to ethanol consumption (18 with no interaction effect), and 53 genes showed an interaction effect (Figure 4.1, Appendix 1). There were 13 genes that were differentially expressed due to running and ethanol, but no interaction. Details about all differentially expressed genes are summarized in Appendix 1. Notable genes include transthyretin (*Ttr*), syntaxin 1b (*Stx1b*), five potassium channels (*Kcnj13*, *Kcnk9*, *Kcne2*, *Kcnj6*, and, *Kcng3*), opioid receptor mu 1 (*Oprm1*), nine genes located in *Etp1* (ethanol preference QTL on chromosome 2), and ten genes located in *Etp5* (ethanol preference QTL on chromosome 9).

We used WebGestalt to test for over-representation of GO, KEGG, and Wikipathways functional groups, using the set of 14207 genes meeting all thresholds as a reference (Table 4.2). The set of ethanol responsive differentially expressed genes was enriched for genes involved in the extracellular matrix (3 genes, $p<0.05$). Differentially expressed genes due to wheel running were primarily involved in transcriptional regulation (17 genes, $p<0.05$) and DNA binding (34 genes, $p<0.05$). Genes differentially expressed due to the interaction of ethanol and wheel running were involved in the extracellular region (12 genes, $p<0.01$) and in complement and coagulation cascades (2 genes, $p<0.05$).

We tested for over-representation of cell-type specific genes, using a database of genes over-expressed in different cell types [112]. There were no over-represented cell-type specific genes in any group of DEGs.

Figure 4.1 Differentially expressed genes, main effects of ethanol and running

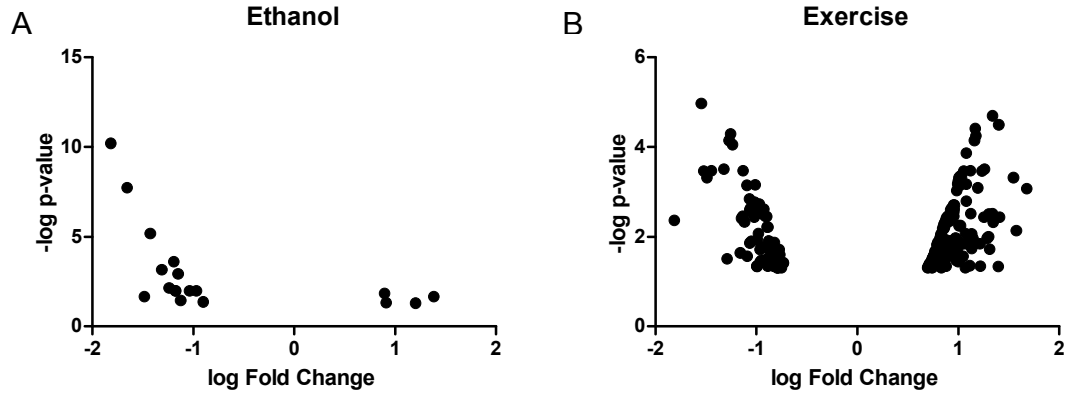

Figure 4.1 shows the distribution of differentially expressed genes with a main effect of access to ethanol (A) and access to running wheel (B). Higher values on the y-axis signify lower p-values. Negative values on the x-axis signify reduced expression due to ethanol or running, positive values indicate increased expression due to ethanol or running. Eighteen genes were differentially expressed due to ethanol, with 14 genes having decreased expression. Two hundred three genes were differentially expressed due to running, 66 down-regulated and 137 up-regulated. P-values have been adjusted for multiple testing.

Table 4.2 Functional over-representation of differentially expressed genes.

Condition[a]	Category Id[b]	Category name[c]	# genes[d]	p-value[e]	FDR[f]
Ethanol	GO:0005604	cellular component:basement membrane	3	0.0001	0.0029
	GO:0044420	cellular component:extracellular matrix part	3	0.0009	0.0087
	GO:0005576	cellular component:extracellular region	6	0.0009	0.0087
	GO:0005578	cellular component:proteinaceous extracellular matrix	3	0.0035	0.0203
	GO:0005581	cellular component:collagen	2	0.0032	0.0203
	GO:0031012	cellular component:extracellular matrix	3	0.005	0.0242
	GO:0044421	cellular component:extracellular region part	4	0.0066	0.0273
	GO:0009887	biological process:organ morphogenesis	5	0.0004	0.0424
Exercise	GO:0003677	molecular function:DNA binding	34	9.11E-05	0.0159
	GO:0001071	molecular function:nucleic acid binding transcription factor activity	17	0.0003	0.0175
	GO:0003700	molecular function:sequence-specific DNA binding transcription factor activity	17	0.0003	0.0175
Interaction	GO:0005576	cellular component:extracellular region	12	7.68E-05	0.005
	WP385	Myometrial Relaxation and Contraction Pathways	4	0.0012	0.0072
	WP449	Complement and Coagulation Cascades	2	0.0057	0.0171
	KEGG:4610	Complement and coagulation cascades	2	0.0064	0.032
	GO:0044421	cellular component:extracellular region part	8	0.0011	0.0357

[a]Genes differentially expressed due to main effect of ethanol, exercise, or ethanol x exercise interaction.

[b]Functional group identifier: GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes, WP (Wikipathways).

[c]Functional group name.

[d]Number of differentially expressed genes in functional group.

[e]Unadjusted p-value.

[f]Benjamini-Hochberg false discovery rate, all significant at FDR < 5%.

4.4.3 WGCNA

EdgeR generated normalized read counts for all 14207 genes meeting minimum thresholds for inclusion. A single WGCNA for all 24 samples produced 29 distinct clusters (modules) of co-expressed genes (Figure 4.2). Each module was named after a color, and no genes were assigned to the 'grey' module, meaning they do not fit a specific co-expression pattern. The number of genes per module ranged from 32 to 2202. Robustness testing confirmed that each module has: greater average adjacency than random, greater average topological overlap than random, and greater scaled intramodular connectivity than scaled extramodular connectivity.

To identify potentially biologically relevant modules, the first principle component of the expression data for each module was calculated using the moduleEigengenes command in R. The calculated module eigengene is representative of the expression pattern of each gene in the module across all samples. A two-way ANOVA (main effects access to ethanol and access to running wheel) suggests that two modules have co-expression patterns responsive to exercise (salmon4, $F_{1,20}$=8.3, p=0.009 and darkslateblue, $F_{1,20}$=8.9, p=0.007) and two modules have co-expression patterns dependent on the interaction of ethanol and exercise (greenyellow, $F_{1,20}$=6.1, p=0.02 and darkorange, $F_{1,20}$=7.4, p=0.01), although none meet the multiple testing threshold of p<0.0017. Module eigengenes and expression patterns for these four modules are shown in Figure 4.3.

Functional over-representation analysis using WebGestalt showed the darkslateblue module was enriched for genes involved in synapse structure (27 genes, p<0.001), voltage-gated ion channels (11 genes, p=0.01), and several signaling

Figure 4.2 Hierarchical clustering of expressed genes, dynamic tree cut, and merged modules.

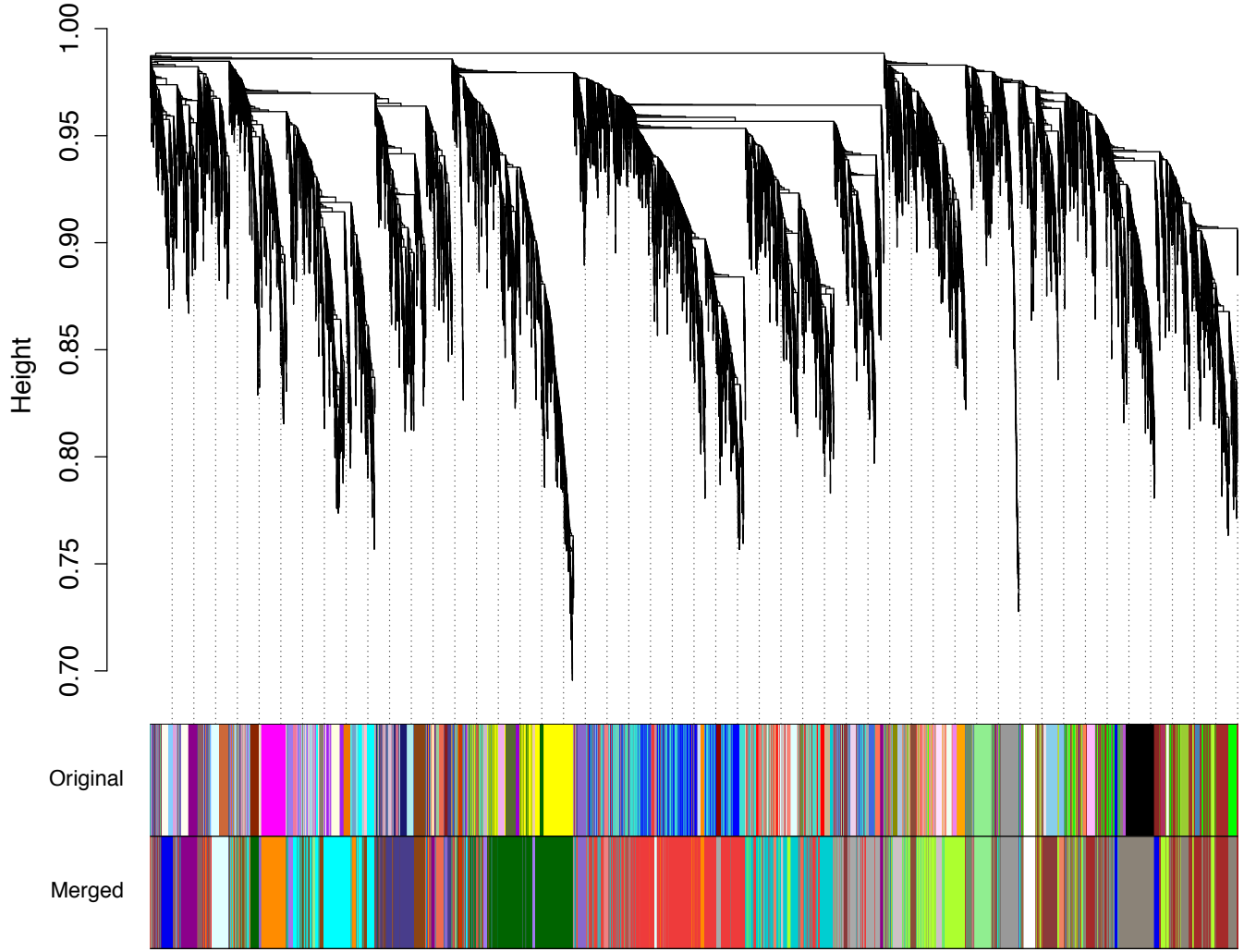

Figure 4.2 shows the results of the hierarchical clustering algorithm and the dynamic tree cut. The y-axis represents a dissimilarity measurement based on topological overlap, with the more similar topological overlaps corresponding to lower heights. Each branch of the dendrogram represents one gene. Branches of the dendrogram are “pruned” into modules, corresponding to each color in the

bottom rows. The top color row shows the module grouping after the initial dynamic tree cut (91 modules), while the bottom color row shows the module grouping after merging similar modules (29 remaining modules).

Figure 4.3 Gene expression within modules and module eigengenes.

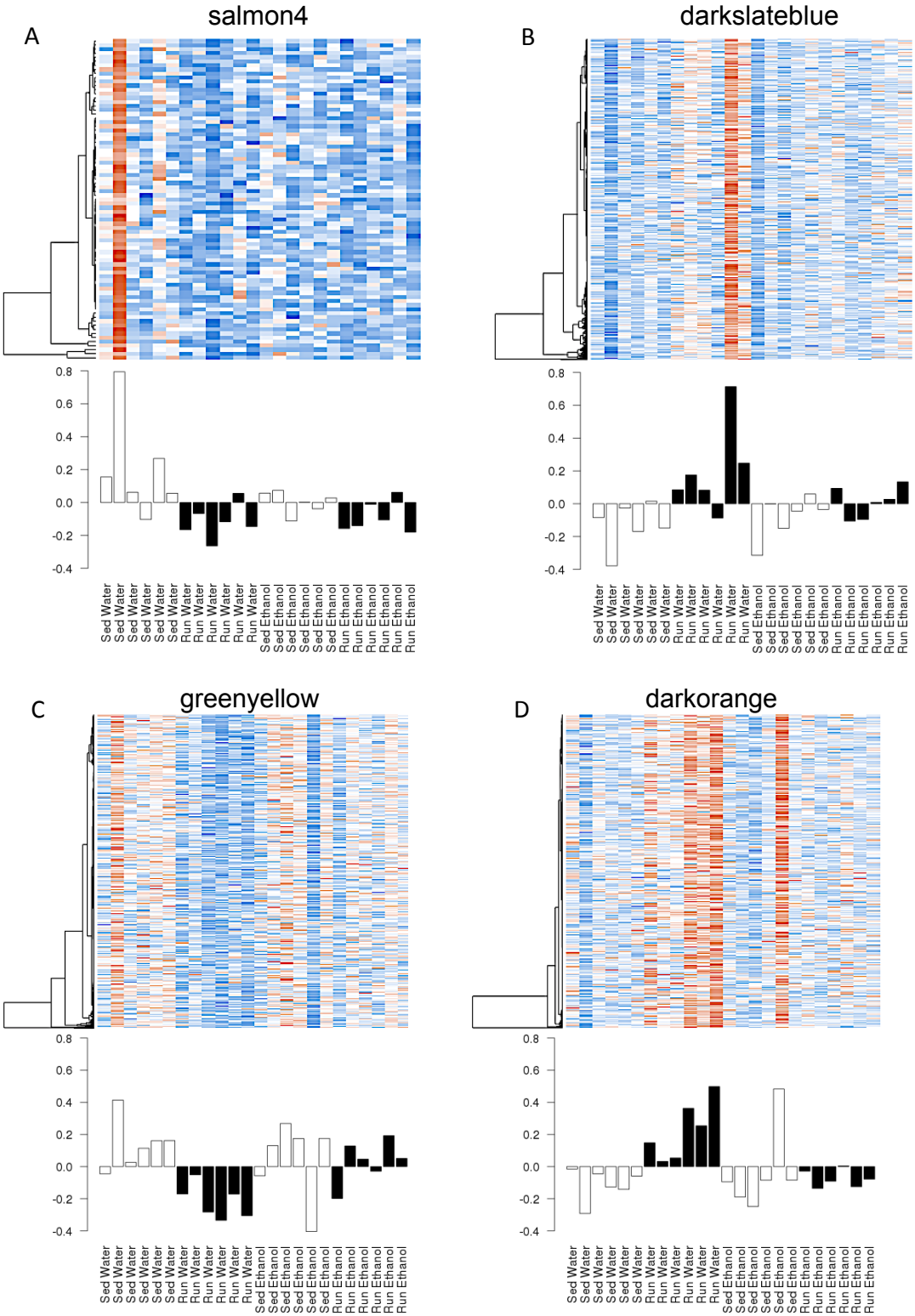

Figure 4.2 shows the four modules corresponding to treatment groups. Heat maps depict expression levels for each gene (rows) in the module, blue colors represent lower expression, red colors represent higher expression. Each column represents one

mouse sample. Barplots of the values of the module eigengene (the first principal component) derived from singular value decomposition are displayed for each module underneath the heatmap. White bars represent sedentary mice, black bars represent running mice. Module eigengenes were tested for relationship to treatment group with two-way ANOVA. Salmon4 (A) and darkslateblue (B) are both running responsive modules, while greenyellow (C) and darkorange (D) both show an interaction of running and ethanol.

pathway including MAPK signaling (8 genes, $p<0.05$) and calcium signaling (6 genes, $p<0.05$). The salmon4 module was enriched for genes involved in cell cycle (2 genes, $p<0.05$). The greenyellow module is enriched for genes involved mRNA processing and transcription regulation (49 genes, $p<0.01$), as well as genes involved in several neurodegenerative diseases: Parkinson's (27 genes, $p<0.001$), Alzheimer's (26 genes, $p=0.001$), and Huntington's (26 genes, $p<0.01$). The darkorange module was enriched in genes involved in axon guidance (11 genes, $p<0.001$), myelin sheath (5 genes, $p<0.01$), regulation of response to stimulus (78 genes, $p<0.001$), regulation of cell communication (74 genes, $p<0.001$), MAPK signaling (13 genes, $p<0.001$), and neurotrophin signaling (10 genes, $p<0.001$). All p-values have been corrected for multiple testing.

There were no cell-type specific genes over-represented in the darkslateblue, salmon4, or greenyellow module. However, the darkorange module was enriched for oligodendrocyte genes (32 genes, hypergeometric $p=0.01$) and trended towards over-representation of astrocyte genes (40 genes, hypergeometric $p=0.06$).

4.5 Discussion

Over the course of 16 days, exercise on a running wheel reduced the amount of ethanol consumed by mice, supporting the hypothesis that under certain conditions, hedonic stimulus from one behavior can substitute for stimulus from another behavior. Transcriptional changes in the striatum, a major component of the mesolimbic dopaminergic reward pathway, could provide insight into the mechanisms underlying reduced ethanol preference.

Table 4.3 Module characteristics and hub genes for significantly associated modules.

Module	# genes[a]	Module eigengene significant[b]	Cell-type enrichment[c]	Hub genes[d]	Locus[e]	Ethanol preference QTL[f]	DEG[g]
salmon4	79	Exercise (p<0.01)		*Cd24a*	chr10:43579169-43584262		Exercise↓
				1700003M02Rik	chr4:34688559-34730206		Exercise↓
				Ccdc33	chr9:58028677-58118823	*Etp5*	Exercise↓
				Kif11	chr19:37376403-37421859		
				Casc1	chr6:145174834-145210970		
darkslateblue	605	Exercise (p<0.01)		*Tet1*	chr10:62804577-62899118		
				Gucy1a2	chr9:3532354-3897342		Exercise↑
				Birc6	chr17:74528295-74703356		
				Sp1	chr15:102406143-102436404		
				Dgki	chr6:36846022-37300184		
greenyellow	1051	Ethanol x Exercise (p<0.05)		*Sqstm1*	chr11:50199366-50210827		
				Fam174a	chr1:95313628-95335284		
				Cd2bp2	chr7:127191660-127196000		
				Snx6	chr12:54746357-54795662		
				Papss1	chr3:131564768-131643670		
darkorange	584	Ethanol x Exercise (p=0.01)	Oligodendrocyte (p=0.01) Astrocyte (p=0.06)	*Jph4*	chr14:55106830-55116935		
				Plekha6	chr1:133246097-133303435		
				Leng8	chr7:4137039-4148173		
				Trim46	chr3:89234177-89246309		
				Crtc2	chr3:90254163-90264125		

[a]Number of co-expressed genes in module.
[b]Module eigengene significantly different due to exercise or ethanol x exercise interaction as determined by two-way ANOVA.
[c]Module genes significantly over-represented by genes previously shown to be cell-type specific [112], as determined by hypergeometric te
[d]Top five hub genes as determined by sorting by module membership.
[e]Genomic location of hub genes.
[f]Hub genes located in ethanol preference QTLs.
[g]Hub genes shown to be differentially expressed genes, and direction of change due to main effect.

Wheel running in rodents induced a myriad of behavioral responses related to tests of stress, anxiety, and depression [83, 139, 183-190]. Furthermore, wheel running was sufficient to reduce voluntary intake of amphetamine [191], cocaine [192, 193], as well as nicotine-seeking during extinction [64, 194]. Taken together with the effects of wheel running on ethanol behaviors, observed across multiple species [58, 59, 61-63, 65-68], these data show a clear effect of hedonic substitution. Understanding the neurobiological components of hedonic substitution will provide for further comprehension of the addiction process, and additional strategies for combating AUDs.

This study validates and extends the findings from Chapter 2, both behaviorally and from expression data. Striatal *Drd1a* showed reduced expression due to access to a running wheel in Chapter 2, and here we see reduced expression due to running wheel, although not quite significant when corrected for multiple testing ($p=0.06$). In addition, we identified 203 additional exercise-responsive DEGs, 18 ethanol-responsive DEGs, and 53 DEGs whose response to ethanol depended on access to a running wheel. Of these 3 sets of DEGs, two possibilities seem most likely for involvement in hedonic substitution. First, genes that regulate ethanol preference may be differentially expressed due to exercise, thereby altering ethanol preference. Second, of the genes that show an interaction effect, some respond to ethanol consumption in the absence of a running wheel—possibly reinforcing the ethanol behavior. These same genes may not respond the same way when the mouse exercises on a running wheel, possibly attenuating the reinforcing effects of ethanol consumption.

Several DEGs are located in previously identified ethanol preference QTLs on chromosomes 2 and 9 [31-37]. A total of 19 DEGs were identified in these regions

(Appendix 1), however, it remains unclear how any of them may specifically affect ethanol preference. Of the 10 DEGs located in *Etp5* on chromosome 9, all were exercise-responsive. Four DEGs in *Etp1* on chromosome 2 showed an interaction effect, the rest were exercise-responsive. No ethanol responsive genes were found in ethanol preference QTL regions.

The sets of DEGs include several genes previously identified as candidate genes for ethanol preference. The expression of opioid receptor mu 1 (*Oprm1*) was increased due to access to a running wheel. The μ-opioid receptor is antagonized by naltrexone, one of the approved pharmacological treatments for AUDs [5]. To speculate, an increase in expression of *Oprm1* suggests possible compensation for a reduction in receptor sensitivity, similar to the effect of antagonizing the receptor.

In mapping an ethanol preference QTL on mouse chromosome 2, syntaxin binding protein 1 (*Stxbp1*) was identified as a candidate gene for ethanol preference [35]. *Stxbp1* was not differentially expressed in our sample, however, syntaxin 1b (*Stx1b*) was higher expressed due to running. These two proteins have been shown to interact, and facilitate neurotransmitter release. *Stx1b* was shown to be more highly expressed in the prefrontal cortex of ethanol preferring P rats compared to non-preferring NP rats [195]. This apparent discrepancy could be due to regional differences in neurotransmitter systems, as striatal neurons are primarily GABA-ergic while cortical neurons have a higher proportion of glutamatergic neurons.

A previous study of gamma-protein kinase c (*Prkcg*) showed the differences in ethanol consumption between null mutant and wildtype mice could be largely correlated with differences in the expression of transthyretin (*Ttr*) [44]. We showed an interaction in

the expression of *Ttr.* Expression was reduced due the ethanol, but that running slightly attenuated this reduction. It was reported that in human alcoholics, transthyretin was reduced immediately after cessation of consumption [196]. This may reflect what was measured in our study, as tissue was collected during the light cycle, when mice typically reduce their ethanol consumption. *Ttr* was also identified as a candidate gene for ethanol sensitivity in Chapter 2 [85], an interesting overlap. Further similarities between the Chapter 2 DEGs and the DEGs in the current study exist. Multiple genes coding for potassium channels were differentially expressed in both studies, again suggesting an overlap between differences in ethanol sensitivity and differences in ethanol consumption.

In addition to differential expression testing, we utilized WGCNA to identify 29 distinct gene co-expression networks. The first principal component of the expression data of each module revealed four modules that related to treatment condition. Two modules, salmon4 and darkslateblue, were exercise-responsive. In general, genes in these modules showed decreased (salmon4) or increased (darkslateblue) expression due to running. Two modules, greenyellow and darkorange, showed an interaction effect between running and ethanol. In the greenyellow module, genes showed decreased expression with running and water, and no decrease or a slight increase with running and ethanol. In the darkorange module, genes showed increased expression with running and water, and a slight decrease in expression with running and ethanol. These two networks, composed of genes with differential responses to exercise depending on whether or not ethanol was consumed, may be more important to hedonic substitution. Interestingly, the darkorange module was enriched for astrocyte and

oligodendrocyte genes, suggesting a role for glial cells in hedonic substitution. In Chapter 2, we proposed a role for glia in determining the sensitivity to ethanol [85]. Again this suggests that one way hedonic substitution might work is through altering ethanol sensitivity.

These data represent the highest resolution transcriptome to date of the striatal transcriptional response to ethanol consumption and wheel running. We identified many exercise-responsive genes that have either been previously implicated in ethanol behaviors (*Oprm1, Ttr,* potassium channels), are associated with previously implicated genes (*Stx1b*), or reside in ethanol preference QTLs. Furthermore, based on gene co-expression networks, we propose a role for glial cells in hedonic substitution. The utility of studying the transcriptome is in the generation of hypotheses for future study. Similarity between groups of genes and gene networks between this study and Chapter 2 lead us to propose that wheel running induces altered sensitivity to ethanol, whereby decreasing ethanol consumption. This would be consistent with previous findings that wheel running does not change ethanol metabolic rates [63] or saccharin consumption (Chapter 2).

Chapter 5

Conclusions

The studies presented in this dissertation were designed to study the influences on ethanol-related phenotypes, both genetic and environmental. A common theme throughout the three studies was the use of measures of transcriptional differences, both transcriptome-wide measures as well as single gene.

In Chapter 2, I examined the baseline transcriptional differences between the Inbred Long Sleep (ILS) and Inbred Short Sleep (ISS) mouse strains. These strains were originally generated through selective breeding to maximize the difference in their sensitivity to ethanol, as a model for the genetic influence on this phenotype. I used multiple bioinformatics resources to identify differentially expressed genes (DEGs), polymorphisms, and co-expression networks that could explain the differences in sensitivity to ethanol. I found multiple genes, *Rassf2, Myo1d, Penk, Pdyn, Ppp1r1b, Ttr,* and 14 potassium channel genes, differentially expressed between strains and each with a plausible reason to be included as candidate genes. In these sets of DEGs, I detected patterns of enrichment for cell-type specific genes, which suggested a role for altered glial/neuronal composition in the striatum of these strains. In addition, I identified multiple variants in *Rassf2* and other genes, many previously unknown that could explain differences in expression and/or functional differences. Finally, using a Weighted Gene Co-expression Network Analysis, I find networks of co-expressed genes, different between strains, involved in several signaling pathways, MAPK, PPAR, and NF-κB.

In Chapter 3, I took a different approach, identifying transcriptional changes in candidate genes in the mesolimbic dopaminergic pathway. Instead of using two strains of mice, I utilized a behavioral model of hedonic substitution to examine the environmental effects of running, ethanol consumption, and both (simultaneous) behaviors on gene expression. First, my results confirmed previous reports of hedonic substitution, and demonstrated that the effect may not translate to the consumption of saccharin. Second, I identified gene expression changes in *Drd1a* in the striatum, *Bdnf* in the hippocampus, and *Slc18a2* in the midbrain. *Bdnf* and *Slc18a2* were especially promising. Running and ethanol consumption had opposing effects on *Bdnf* expression, and there was an interaction effect on *Slc18a2* expression. These results highlighted the complexity of ethanol behaviors as polygenic, and involving multiple brain regions.

In Chapter 4, using the same behavioral model as in the previous chapter, I sequenced the striatal transcriptome to identify additional candidate genes for hedonic substitution. I found multiple DEGs, *Oprm1*, *Ttr*, *Stx1b*, and several potassium channel genes. I also utilized WGCNA to identify co-expression networks, implicating astrocytes and oligodendrocytes in hedonic substitution. These results, and the similarity between them and the DEGs and functional groups from chapter 2, lead me to propose that one potential mechanism for hedonic substitution is the alteration of sensitivity to ethanol.

Alcohol use disorders (AUDs) are complex diseases with an etiology that encompasses multiple genes and gene networks across many brain regions, as well as environmental input. The use of transcriptome-wide analysis techniques has been productive in identifying new candidate genes and new directions for research in this area. These findings will spur further research into the etiology of AUDs, the

development of pharmacological interventions, and provide a framework for understanding the role of hedonic substitution as a behavioral intervention.

References

1. WHO | World Health Organization. www.who.int

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

3. Warren KR, Hewitt BG: NIAAA: Advancing Alcohol Research for 40 Years. *Alcohol Research & Health* 2010:1–14.

4. National Institute for Alcohol Abuse and Alcoholism (NIAAA). www.niaaa.nih.gov

5. Johnson BA: Medication treatment of different types of alcoholism. *Am J Psychiatry* 2010, 167:630–639.

6. Gunzerath L, Hewitt BG, Li T-K, Warren KR: Alcohol research: past, present, and future. *Ann N Y Acad Sci* 2011, 1216:1–23.

7. Hart HH: Personality factors in alcoholism. *Archives of neurology and psychiatry* 1930, 24:116–134.

8. Cloninger CR, Bohman M, Sigvardson S: Inheritance of Alcohol-Abuse - Cross-Fostering Analysis of Adopted Men. *Arch Gen Psychiatry* 1981, 38:861–868.

9. Cotton NSN: The familial incidence of alcoholism: a review. *J Stud Alcohol* 1979, 40:89–116.

10. Dick DM, Prescott C, McGue M: The Genetics of Substance Use and Substance Use Disorders. *Handbook of Behavior Genetics* 2009:433–453.

11. Foroud T, Edenberg HJ, Crabbe JC: Genetic Research: Who is at risk for alcoholism? *Alcohol Research & Health* 2010:1–13.

12. Dick DM, Foroud T: Genetic strategies to detect genes involved in alcoholism and alcohol-related traits. *Alcohol Res Health* 2002, 26:172–180.

13. Edenberg HJ, Foroud T: The genetics of alcoholism: identifying specific genes through family studies. *Addict Biol* 2006, 11:386–396.

14. Xuei X, Flury-Wetherill L, Dick DM, Goate A, Tischfield J, Nurnberger J, Schuckit MA, Kramer J, Kuperman S, Hesselbrock V, Porjesz B, Foroud T, Edenberg HJ: GABRR1 and GABRR2, encoding the GABA-A receptor subunits rho1 and rho2, are associated with alcohol dependence. *Am J Med Genet B Neuropsychiatr Genet* 2010, 153B:418–427.

15. Kareken DA, Liang T, Wetherill L, Dzemidzic M, Bragulat V, Cox C, Talavage T, O'Connor SJ, Foroud T: A polymorphism in GABRA2 is associated with the medial

frontal response to alcohol cues in an fMRI study. *Alcohol Clin Exp Res* 2010, 34:2169–2178.

16. Dick DM, Plunkett J, Wetherill LF, Xuei X, Goate A, Hesselbrock V, Schuckit MA, Crowe RR, Edenberg HJ, Foroud T: Association between GABRA1 and drinking behaviors in the collaborative study on the genetics of alcoholism sample. *Alcohol Clin Exp Res* 2006, 30:1101–1110.

17. Agrawal A, Edenberg HJ, Foroud T, Bierut LJ, Dunne G, Hinrichs AL, Nurnberger JI, Crowe R, Kuperman S, Schuckit MA, Begleiter H, Porjesz B, Dick DM: Association of GABRA2 with drug dependence in the collaborative study of the genetics of alcoholism sample. *Behavior Genetics* 2006, 36:640–650.

18. Edenberg HJH, Dick DMD, Xuei XX, Tian HH, Almasy LL, Bauer LOL, Crowe RRR, Goate AA, Hesselbrock VV, Jones KK, Kwon JJ, Li T-KT, Nurnberger JIJ, O'Connor SJS, Reich TT, Rice JJ, Schuckit MAM, Porjesz BB, Foroud TT, Begleiter HH: Variations in GABRA2, Encoding the α2 Subunit of the GABA-A Receptor, Are Associated with Alcohol Dependence and with Brain Oscillations. *The American Journal of Human Genetics* 2004, 74:10–10.

19. Wang JC, Hinrichs AL, Bertelsen S, Stock H, Budde JP, Dick DM, Bucholz KK, Rice J, Saccone NL, Edenberg HJ, Hesselbrock V, Kuperman S, Schuckit MA, Bierut LJ, Goate AM: Functional variants in TAS2R38 and TAS2R16 influence alcohol consumption in high-risk families of African-American origin. *Alcohol Clin Exp Res* 2007, 31:209–215.

20. Wetherill LF, Schuckit MA, Hesselbrock V, Xuei X, Liang T, Dick DM, Kramer J, Nurnberger JI, Tischfield JA, Porjesz B, Edenberg HJ, Foroud T: Neuropeptide Y receptor genes are associated with alcohol dependence, alcohol withdrawal phenotypes, and cocaine dependence. *Alcohol Clin Exp Res* 2008, 32:2031–2040.

21. Edenberg HJ, Xuei X, Wetherill LF, Bierut LJ, Bucholz KK, Dick DM, Hesselbrock V, Kuperman S, Porjesz B, Schuckit MA, Tischfield JA, Almasy LA, Nurnberger JI, Foroud T: Association of NFKB1, which encodes a subunit of the transcription factor NF-kappaB, with alcohol dependence. *Hum Mol Genet* 2008, 17:963–970.

22. Edenberg HJ: The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health* 2007, 30:5–13.

23. Edenberg HJ, Xuei X, Chen H-J, Tian H, Wetherill LF, Dick DM, Almasy L, Bierut L, Bucholz KK, Goate A, Hesselbrock V, Kuperman S, Nurnberger J, Porjesz B, Rice J, Schuckit M, Tischfield J, Begleiter H, Foroud T: Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Hum Mol Genet* 2006, 15:1539–1549.

24. Tu GCG, Israel YY: Alcohol consumption by orientals in North America is predicted largely by a single gene. *Behavior Genetics* 1995, 25:59–65.

25. Yoneyama N, Crabbe JC, Ford M, Murillo A, Finn DA: Voluntary ethanol consumption in 22 inbred mouse strains. *Alcohol* 2008, 42:149–160.

26. Belknap JK, Crabbe JC, Young ER: Voluntary consumption of ethanol in 15 inbred mouse strains. *Psychopharmacology (Berl)* 1993, 112:503–510.

27. McClearn GE, Rodgers DA: Differences in alcohol preference among inbred strains of mice. *QJ Stud Alcohol* 1959, 20:691–695.

28. Kakihana R, Brown DR, McClearn GE, Tabershaw IR: Brain sensitivity to alcohol in inbred mouse strains. *Science* 1966, 154:1574–1575.

29. Belknap JK, MacInnes JW, McClearn GE: Ethanol sleep times and hepatic alcohol and aldehyde dehydrogenase activities in mice. *Physiology & Behavior* 1972, 9:453–457.

30. Grisel JEJ, Metten PP, Wenger CDC, Merrill CMC, Crabbe JCJ: Mapping of quantitative trait loci underlying ethanol metabolism in BXD recombinant inbred mouse strains. *Alcohol Clin Exp Res* 2002, 26:610–616.

31. Belknap JK, Atkins AL: The replicability of QTLs for murine alcohol preference drinking behavior across eight independent studies. *Mamm Genome* 2001, 12:893–899.

32. Phillips TJ, Belknap JK, Buck KJ, Cunningham CL: Genes on mouse Chromosomes 2 and 9 determine variation in ethanol consumption. *Mamm Genome* 1998, 9:936–941.

33. Phillips TJ, Crabbe JC, Metten P, Belknap JK: Localization of Genes Affecting Alcohol-Drinking in Mice. *Alcohol Clin Exp Res* 1994, 18:931–941.

34. Belknap JK, Richards SP, OToole LA, Helms ML: Short-term selective breeding as a tool for QTL mapping: Ethanol preference drinking in mice. *Behavior Genetics* 1997, 27:55–66.

35. Fehr C, Shirley RL, Crabbe JC, Belknap JK, Buck KJ, Phillips TJ: The syntaxin binding protein 1 gene (Stxbp1) is a candidate for an ethanol preference drinking locus on mouse chromosome 2. *Alcohol Clin Exp Res* 2005, 29:708–720.

36. Hitzemann RJ, Reed C, Malmanger B, Lawler M, Hitzemann B, Cunningham B, McWeeney S, Belknap JK, Harrington C, Buck KJ, Phillips TJ, Crabbe JC: On the integration of alcohol-related quantitative trait loci and gene expression analyses. *Alcohol Clin Exp Res* 2004, 28:1437–1448.

37. Crabbe JC, Phillips TJ, Belknap JK: The Complexity of Alcohol Drinking: Studies in Rodent Genetic Models. *Behavior Genetics* 2010, 40:737–750.

38. Schwab SG, Franke PE, Hoefgen B, Guttenthaler V, Lichtermann D, Trixler M, Knapp M, Maier W, Wildenauer DB: Association of DNA Polymorphisms in the Synaptic Vesicular Amine Transporter Gene (SLC18A2) with Alcohol and Nicotine Dependence.

Neuropsychopharmacology 2005, 30:2263–2268.

39. Thanos PK, Volkow ND, Freimuth P, Umegaki H, Ikari H, Roth G, Ingram DK, Hitzemann RJ: Overexpression of dopamine D2 receptors reduces alcohol self-administration. *J Neurochem* 2001, 78:1094–1103.

40. Thanos PK, Taintor NB, Rivera SN, Umegaki H, Ikari H, Roth G, Ingram DK, Hitzemann RJ, Fowler JS, Gatley SJ, Wang G-J, Volkow ND: DRD2 Gene Transfer Into the Nucleus Accumbens Core of the Alcohol Preferring and Nonpreferring Rats Attenuates Alcohol Drinking. *Alcohol Clin Exp Res* 2006, 28:720–728.

41. Bulwa ZB, Sharlin JA, Clark PJ, Bhattacharya TK, Kilby CN, Wang Y, Rhodes JS: Increased consumption of ethanol and sugar water in mice lacking the dopamine D2 long receptor. *Alcohol* 2011, 45:631–639.

42. Blednov YA, Walker D, Martinez M, Harris RA: Reduced alcohol consumption in mice lacking preprodynorphin. *Alcohol* 2006, 40:73–86.

43. Hall F, Sora I, Uhl G: Sex-dependent modulation of ethanol consumption in vesicular monoamine transporter 2 (VMAT2) and dopamine transporter (DAT) knockout mice. *Neuropsychopharmacology* 2003, 28:620–628.

44. Smith AM, Bowers BJ, Radcliffe RA, Wehner JM: Microarray analysis of the effects of a gamma-protein kinase C null mutation on gene expression in striatum: a role for transthyretin in mutant phenotypes. *Behavior Genetics* 2006, 36:869–881.

45. Schuckit MA: Self-rating of alcohol intoxication by young men with and without family histories of alcoholism. *J Stud Alcohol* 1980, 41:242–249.

46. Schuckit MA: Low level of response to alcohol as a predictor of future alcoholism. *American Journal of Psychiatry* 1994, 151:184–189.

47. Schuckit MA, Smith TL: The relationships of a family history of alcohol dependence, a low level of response to alcohol and six domains of life functioning to the development of alcohol use disorders. *J Stud Alcohol* 2000, 61:827–835.

48. Schuckit MA, Tsuang JW, Anthenelli RM, Tipp JE, Nurnberger JI: Alcohol challenges in young men from alcoholic pedigrees and control families: a report from the COGA project. *J Stud Alcohol* 1996, 57:368–377.

49. Schuckit MA: An overview of genetic influences in alcoholism. *J Subst Abuse Treat* 2009, 36:S5–14.

50. McClearn G, Kakihana R: Selective Breeding for Ethanol Sensitivity: Short-Sleep and Long-Sleep Mice. In *Development of Animal Models as Pharmacogenetic Tools*. Edited by McClearn G, Deitrich RA, Erwin V. Washington, DC: US Government Printing Office, DHHS Publication No. [ADM] 81-113; 1981:147–159.

51. Saba LM, Bennett B, Hoffman PL, Barcomb K, Ishii T, Kechris K, Tabakoff B: A systems genetic analysis of alcohol drinking by mice, rats and men: influence of brain GABAergic transmission. *Neuropharmacology* 2011, 60:1269–1280.

52. Markel PD, Fulker DW, Bennett B, Corley RP, DeFries JC, Erwin VG, Johnson TE: Quantitative trait loci for ethanol sensitivity in the LS x SS recombinant inbred strains: interval mapping. *Behavior Genetics* 1996, 26:447–458.

53. Markel PD, Bennett B, Beeson M, Gordon LG, Johnson TE: Confirmation of quantitative trait loci for ethanol sensitivity in long-sleep and short-sleep mice. *Genome Research* 1997, 7:92–99.

54. Bennett B, Carosone-Link P, Zahniser NR, Johnson TE: Confirmation and fine mapping of ethanol sensitivity quantitative trait loci, and candidate gene testing in the LXS recombinant inbred mice. *Journal of Pharmacology and Experimental Therapeutics* 2006, 319:299–307.

55. Bennett B, Beeson M, Gordon LG, Carosone-Link P, Johnson TE: Genetic dissection of quantitative trait loci specifying sedative/hypnotic sensitivity to ethanol: mapping with interval-specific congenic recombinant lines. *Alcohol Clin Exp Res* 2002, 26:1615–1624.

56. Bennett B, Carosone-Link P, Beeson M, Gordon LG, Phares-Zook N, Johnson TE: Genetic dissection of quantitative trait locus for ethanol sensitivity in long- and short-sleep mice. *Genes Brain Behav* 2008, 7:659–668.

57. Maclaren EJ, Bennett B, Johnson TE, Sikela JM: Expression profiling identifies novel candidate genes for ethanol sensitivity QTLs. *Mamm Genome* 2006, 17:147–156.

58. McMillan DE, McClure GY, Hardwick WC: Effects of access to a running wheel on food, water and ethanol intake in rats bred to accept ethanol. *Drug and Alcohol Dependence* 1995, 40:1–7.

59. McMillan DE: Effects of access to a running wheel on ethanol intake in rats under schedule-induced polydipsia. *Currents in alcohol* 1978, 3:221–235.

60. Werme M, Lindholm S, Thoren P, Franck J, Brene S: Running increases ethanol preference. *Behavioural Brain Research* 2002, 133:301–308.

61. Ozburn AR, Harris RA, Blednov YA: Wheel running, voluntary ethanol consumption, and hedonic substitution. *Alcohol* 2008, 42:417–424.

62. Hammer SB, Ruby CL, Brager AJ, Prosser RA, Glass JD: Environmental Modulation of Alcohol Intake in Hamsters: Effects of Wheel Running and Constant Light Exposure. *Alcohol Clin Exp Res* 2010, 34:1–8.

63. Ehringer MA, Hoft NR, Zunhammer M: Reduced alcohol consumption in mice with access to a running wheel. *Alcohol* 2009, 43:443–452.

64. Smith MA, Lynch WJ: Exercise as a Potential Treatment for Drug Abuse: Evidence from Preclinical Studies. *Frontiers in Psychiatry* 2012, 2.

65. Weinstock J: A review of exercise as intervention for sedentary hazardous drinking college students: rationale and issues. *J Am Coll Health* 2010, 58:539–544.

66. Murphy TJ, Pagano RR, Marlatt GA: Lifestyle modification with heavy alcohol drinkers: effects of aerobic exercise and meditation. *Addict Behav* 1986, 11:175–186.

67. Correia CJ, Benson TA, Carey KB: Decreased substance use following increases in alternative behaviors: a preliminary investigation. *Addict Behav* 2005, 30:19–27.

68. Werch CE, Bian H, Carlson JM, Moore MJ, Diclemente CC, Huang I-C, Ames SC, Thombs D, Weiler RM, Pokorny SB: Brief integrative multiple behavior intervention effects and mediators for adolescents. *Journal of behavioral medicine* 2010.

69. Gutgesell ME, Timmerman M, Keller A: Reported alcohol use and behavior in long-distance runners. *Medicine and science in sports and exercise* 1996, 28:1063–1070.

70. Ussher MH, Sampuran AK, Doshi R, West R, Drummond DC: Acute effect of a brief bout of exercise on alcohol urges. *Addiction* 2004, 99:1542–1547.

71. Xu Y, Ehringer MA, Yang F, Sikela JM: Comparison of global brain gene expression profiles between inbred long-sleep and inbred short-sleep mice by high-density gene array hybridization. *Alcohol Clin Exp Res* 2001, 25:810–818.

72. Maclaren EJ, Sikela JM: Cerebellar gene expression profiling and eQTL analysis in inbred mouse strains selected for ethanol sensitivity. *Alcohol Clin Exp Res* 2005, 29:1568–1579.

73. Radcliffe RA, Lee MJ, Williams RW: Prediction of cis-QTLs in a pair of inbred mouse strains with the use of expression and haplotype data from public databases. *Mamm Genome* 2006, 17:629–642.

74. Treadwell JA, Singh SM: Microarray analysis of mouse brain gene expression following acute ethanol treatment. *Neurochemical research* 2004, 29:357–369.

75. Kerns RT, Ravindranathan A, Hassan S, Cage MP, York T, Sikela JM, Williams RW, Miles MF: Ethanol-responsive brain region expression networks: implications for behavioral responses to acute ethanol in DBA/2J versus C57BL/6J mice. *Journal of Neuroscience* 2005, 25:2255–2266.

76. Saito M, Smiley J, Toth R, Vadasz C: Microarray analysis of gene expression in rat hippocampus after chronic ethanol treatment. *Neurochemical research* 2002, 27:1221–1229.

77. Saito M, Szakall I, Toth R, Kovacs KM, Oros M, Prasad VVTS, Blumenberg M, Vadasz C: Mouse striatal transcriptome analysis: effects of oral self-administration of

alcohol. *Alcohol* 2004, 32:223–241.

78. Bell RL, Kimpel MW, McClintick JN, Strother WN, Carr LG, Liang T, Rodd ZA, Mayfield RD, Edenberg HJ, McBride WJ: Gene expression changes in the nucleus accumbens of alcohol-preferring rats following chronic ethanol consumption. *Pharmacol Biochem Behav* 2009, 94:131–147.

79. McBride WJ, Kimpel MW, Schultz JA, McClintick JN, Edenberg HJ, Bell RL: Changes in gene expression in regions of the extended amygdala of alcohol-preferring rats after binge-like alcohol drinking. *Alcohol* 2010, 44:171–183.

80. Mulligan MK, Rhodes JS, Crabbe JC, Mayfield RD, Harris RA, Ponomarev I: Molecular Profiles of Drinking Alcohol to Intoxication in C57BL/6J Mice. *Alcohol Clin Exp Res* 2011, 35:1–12.

81. Tong L, Shen H, Perreau VM, Balazs R, Cotman CW: Effects of exercise on gene-expression profile in the rat hippocampus. *Neurobiol Dis* 2001, 8:1046–1056.

82. Greenwood BN, Foley TE, Le TV, Strong PV, Loughridge AB, Day HEW, Fleshner M: Long-term voluntary wheel running is rewarding and produces plasticity in the mesolimbic reward pathway. *Behavioural Brain Research* 2011, 217:354–362.

83. Dishman RK, Berthoud H-R, Booth FW, Cotman CW, Edgerton VR, Fleshner MR, Gandevia SC, Gomez-Pinilla F, Greenwood BN, Hillman CH, Kramer AF, Levin BE, Moran TH, Russo-Neustadt AA, Salamone JD, Van Hoomissen JD, Wade CE, York DA, Zigmond MJ: Neurobiology of exercise. *Obesity (Silver Spring)* 2006, 14:345–356.

84. Foley TE, Fleshner MR: Neuroplasticity of dopamine circuits after exercise: implications for central fatigue. *Neuromolecular Med* 2008, 10:67–80.

85. Darlington TM, Ehringer MA, Larson C, Phang TL, Radcliffe RA: Transcriptome analysis of Inbred Long Sleep and Inbred Short Sleep mice. *Genes Brain Behav* 2013, 12:263–274.

86. Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009, 10:57–63.

87. Marguerat S, Bähler J: RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences* 2010, 67:569–579.

88. Iancu OD, Kawane S, Bottomly D, Searles R, Hitzemann RJ, McWeeney S: Utilizing RNA-Seq data for de novo coexpression network inference. *Bioinformatics* 2012, 28:1592–1597.

89. DeFries JC, Wilson JR, Erwin VG, Petersen DR: LS X SS recombinant inbred strains of mice: initial characterization. *Alcohol Clin Exp Res* 1989, 13:196–200.

90. Erwin VG, Jones BC, Radcliffe RA: Further characterization of LSxSS recombinant

inbred strains of mice: activating and hypothermic effects of ethanol. *Alcohol Clin Exp Res* 1990, 14:200–204.

91. Gehle V, Erwin V: The genetics of acute functional tolerance and initial sensitivity to ethanol for an ataxia test in the LSxSS RI strains. *Alcohol Clin Exp Res* 2000, 24:579–587.

92. Christensen SC, Johnson TE, Markel PD, Clark VJ, Fulker DW, Corley RP, Collins AC, Wehner JM: Quantitative Trait Locus Analyses of Sleep-Times Induced by Sedative-Hypnotics in LSXSS Recombinant Inbred Strains of Mice. *Alcohol Clin Exp Res* 1996, 20:543–550.

93. Ehringer MA, Thompson J, Conroy O, Xu Y, Yang F, Canniff J, Beeson M, Gordon LG, Bennett B, Johnson TE, Sikela JM: High-throughput sequence identification of gene coding variants within alcohol-related QTLs. *Mamm Genome* 2001, 12:657–663.

94. Shendure J, Ji H: Next-generation DNA sequencing. *Nat Biotechnol* 2008, 26:1–11.

95. Trapnell C, Pachter L, Salzberg SL: TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009, 25:1105–1111.

96. Trapnell C, Salzberg SL: How to map billions of short reads onto genomes. *Computation Biology* 2009, 27:1–3.

97. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: Integrative genomics viewer. *Nat Biotechnol* 2011, 29:24–26.

98. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010, 28:511–515.

99. Brooks MJ, Rajasimha HK, Roger JE, Swaroop A: Next-generation sequencing facilitates quantitative analysis of wild-type and Nrl−/− retinal transcriptomes. *Molecular Vision* 2011, 17:3034.

100. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis CA, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, Sturgill D, Tuch BB, Zaleski C, Zhang D, Blanchette M, Dudoit S, Eads B, Green RE, Hammonds A, Jiang L, Kapranov P, et al.: The developmental transcriptome of Drosophila melanogaster. *Nature* 2010, 471:473–479.

101. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics (BMCBI) 9* 2008.

102. Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S: Weighted Gene Coexpression Network Analysis: State of the Art. *J Biopharm Stat* 2010, 20:281–300.

103. Langfelder P, Horvath S: Fast R Functions for Robust Correlations and Hierarchical Clustering. *J Stat Softw* 2012, 46:–.

104. Benjamini Y, Hochberg Y: Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B …* 1995, 57:289–300.

105. Duncan D, Prodduturi N, Zhang B: WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. *BMC Bioinformatics* 2010, 11:P10.

106. Zhang B, Kirov S, Snoddy JR: WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Research* 2005, 33(Web Server issue):W741–8.

107. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000, 25:25–29.

108. Kanehisa MM, Goto SS: KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 2000, 28:27–30.

109. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* 2011, 40:D109–D114.

110. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, Pico AR: WikiPathways: building research communities on biological pathways. *Nucleic Acids Research* 2012, 40(Database issue):D1301–7.

111. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C: WikiPathways: pathway editing for the people. *Plos Biol* 2008, 6:e184–1407.

112. Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA, Thompson WJ, Barres BA: A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *Journal of Neuroscience* 2008, 28:264–278.

113. Kent WJ: BLAT---The BLAST-Like Alignment Tool. *Genome Research* 2002, 12:656–664.

114. Walter NA, McWeeney SK, Peters ST, Belknap JK, Hitzemann RR, Buck KJ: SNPs matter: impact on detection of differential expression. *Nat Meth* 2007, 4:679–680.

115. Donninger H, Hesson L, Vos M, Beebe K, Gordon L, Sidransky D, Liu JW, Schlegel T, Payne S, Hartmann A, Latif F, Clark GJ: The Ras effector RASSF2 controls

the PAR-4 tumor suppressor. *Mol Cell Biol* 2010, 30:2608–2620.

116. Benesh AE, Fleming JT, Chiang C, Carter BD, Tyska MJ: Expression and localization of myosin-1d in the developing nervous system. *Brain Res* 2012, 1440:9–22.

117. Nylander I, Hyytia P, Forsander O, Terenius L: Differences between Alcohol-Preferring (AA) and Alcohol-Avoiding (ANA) Rats in the Prodynorphin and Proenkephalin Systems. *Alcohol Clin Exp Res* 1994, 18:1272–1279.

118. Brick J, Horowitz GP: Tolerance and cross-tolerance to morphine and ethanol in mice selectively bred for differential sensitivity to ethanol. *J Stud Alcohol* 1983, 44:770–779.

119. Yger M, Girault J-A: DARPP-32, jack of all trades... master of which? *Front Behav Neurosci* 2011, 5:1–14.

120. Harris RA, McQuilkin SJ, Paylor R, Abeliovich A, Tonegawa S, Wehner JM: Mutant mice lacking the gamma isoform of protein kinase C show decreased behavioral actions of ethanol and altered function of gamma-aminobutyrate type A receptors. *Proc Natl Acad Sci USA* 1995, 92:3658–3662.

121. Bowers BJ, Wehner JM: Ethanol consumption and behavioral impulsivity are increased in protein kinase C gamma null mutant mice. *J Neurosci* 2001, 21:RC180–RC180.

122. Brodie MS, Pesold C, Appel SB: Ethanol Directly Excites Dopaminergic Ventral Tegmental Area Reward Neurons. *Alcohol Clin Exp Res* 1999, 23:1848–1852.

123. Hopf FWF, Bowers MSM, Chang S-JS, Chen BTB, Martin MM, Seif TT, Cho SLS, Tye KK, Bonci AA: Reduced Nucleus Accumbens SK Channel Activity Enhances Alcohol Seeking during Abstinence. *Neuron* 2010, 65:13–13.

124. Hopf FW, Simms JA, Chang S-J, Seif T, Bartlett SE, Bonci A: Chlorzoxazone, an SK-Type Potassium Channel Activator Used in Humans, Reduces Excessive Alcohol Intake in Rats. *Biological Psychiatry* 2011, 69:618–624.

125. Mulholland PJP: K(Ca)2 channels: Novel therapeutic targets for treating alcohol withdrawal and escalation of alcohol consumption. *Alcohol* 2012, 46:309–315.

126. Bottomly D, Walter NAR, Hunter JE, Darakjian P, Kawane S, Buck KJ, Searles RP, Mooney M, McWeeney SK, Hitzemann RJ: Evaluating Gene Expression in C57BL/6J and DBA/2J Mouse Striatum Using RNA-Seq and Microarrays. *PLoS ONE* 2011, 6:e17820.

127. Carnicella S, Kharazia V, Jeanblanc J, Janak PH, Ron D: GDNF is a fast-acting potent inhibitor of alcohol consumption and relapse. *Proceedings of the National Academy of Sciences* 2008, 105:8114–8119.

128. Foroud T, Li TK: Genetics of alcoholism: a review of recent studies in human and animal models. *Am J Addict* 1999, 8:261–278.

129. Koob GF, Volkow ND: Neurocircuitry of Addiction. *Neuropsychopharmacology* 2009:1–22.

130. Harris RA, Trudell JR, Mihic SJ: Ethanol's molecular targets. *Science signaling* 2008, 1:re7.

131. Nestler EJ: Is there a common molecular pathway for addiction? *Nat Neurosci* 2005, 8:1445–1449.

132. McMillan DE: Abstract-Effects of access to a running wheel on ethanol intake in rats under schedule-induced polydipsia. *Alcohol Clin Exp Res* 1977, 1:160–160.

133. Di Chiara G, Imperato A: Preferential stimulation of dopamine release in the nucleus accumbens by opiates, alcohol, and barbiturates: studies with transcerebral dialysis in freely moving rats. *Ann N Y Acad Sci* 1986, 473:367–381.

134. Di Chiara G, Imperato A: Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. *Proc Natl Acad Sci USA* 1988, 85:5274–5278.

135. Di Chiara G, Imperato A: Ethanol preferentially stimulates dopamine release in the nucleus accumbens of freely moving rats. *Eur J Pharmacol* 1985, 115:131–132.

136. Hoffmann PL, Tabakoff B: Adaptive changes in the dopamine system produced by chronic ethanol feeding. *Drug and Alcohol Dependence* 1979, 4:255–260.

137. Chaouloff F: Physical exercise and brain monoamines: a review. *Acta Physiologica Scandinavica* 1989, 137:1–13.

138. Gayer GG, Gordon A, Miles MF: Ethanol increases tyrosine hydroxylase gene expression in N1E-115 neuroblastoma cells. *J Biol Chem* 1991, 266:22279–22284.

139. Dishman RK: Brain monoamines, exercise, and behavioral stress: animal models. *Medicine and science in sports and exercise* 1997, 29:63–74.

140. Szot P, White S, Veith R, Rasmussen D: Reduced gene expression for dopamine biosynthesis and transport in midbrain neurons of adult male rats exposed prenatally to ethanol. *Alcohol Clin Exp Res* 1999, 23:1643–1649.

141. Oliva JMJ, Manzanares JJ: Gene transcription alterations associated with decrease of ethanol intake induced by naltrexone in the brain of Wistar rats. *Neuropsychopharmacology* 2007, 32:1358–1369.

142. Knab AM, Bowen RS, Hamilton AT, Gulledge AA, Lightfoot JT: Altered dopaminergic profiles: implications for the regulation of voluntary physical activity.

Behavioural Brain Research 2009, 204:147–152.

143. Nascimento PSD, Lovatel GA, Barbosa S, Ilha J, Centenaro LA, Malysz T, Xavier LL, Schaan BD, Achaval M: Treadmill training improves motor skills and increases tyrosine hydroxylase immunoreactivity in the substantia nigra pars compacta in diabetic rats. *Brain Res* 2011, 1382:173–180.

144. 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:e21100–.

145. Lin Z: SLC18A2 promoter haplotypes and identification of a novel protective factor against alcoholism. *Hum Mol Genet* 2005, 14:1393–1404.

146. Schwab SG, Franke PE, Hoefgen B, Guttenthaler V, Lichtermann D, Trixler M, Knapp M, Maier W, Wildenauer DB: Association of DNA polymorphisms in the synaptic vesicular amine transporter gene (SLC18A2) with alcohol and nicotine dependence. *Neuropsychopharmacology* 2005, 30:2263–2268.

147. Eiden LE, Weihe E: VMAT2: a dynamic regulator of brain monoaminergic neuronal function interacting with drugs of abuse. *Ann N Y Acad Sci* 2011, 1216:86–98.

148. Savelieva KV, Caudle WM, Miller GW: Altered ethanol-associated behaviors in vesicular monoamine transporter heterozygote knockout mice. *Alcohol* 2006, 40:87–94.

149. Mathews TA, John CE, Lapa GB, Budygin EA, Jones SR: No role of the dopamine transporter in acute ethanol effects on striatal dopamine dynamics. *Synapse* 2006, 60:288–294.

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

151. Schacht JP, Anton RF, Voronin KE, Randall PK, Li X, Henderson S, Myrick H: Interacting Effects of Naltrexone and OPRM1 and DAT1 Variation on the Neural Response to Alcohol Cues. *Neuropsychopharmacology* 2013, 38:414–422.

152. Savelieva KVK, Caudle WMW, Findlay GSG, Caron MGM, Miller GWG: Decreased ethanol preference and consumption in dopamine transporter female knock-out mice. *Alcohol Clin Exp Res* 2002, 26:758–764.

153. Bice PJ, Liang T, Zhang L, Strother WN, Carr LG: Drd2 expression in the high alcohol-preferring and low alcohol-preferring mice. *Mamm Genome* 2008, 19:69–76.

154. Dick DM, Wang JC, Plunkett J, Aliev F, Hinrichs A, Bertelsen S, Budde JP, Goldstein EL, Kaplan D, Edenberg HJ, Nurnberger J, Hesselbrock V, Schuckit MA, Kuperman S, Tischfield J, Porjesz B, Begleiter H, Bierut LJ, Goate A: Family-based association analyses of alcohol dependence phenotypes across DRD2 and neighboring

gene ANKK1. *Alcohol Clin Exp Res* 2007, 31:1645–1653.

155. Mathes WF, Nehrenberg DL, Gordon R, Hua K, Garland T Jr., Pomp D: Dopaminergic dysregulation in mice selectively bred for excessive exercise or obesity. *Behavioural Brain Research* 2010, 210:155–163.

156. Contet C, Gardon O, Filliol D, Becker JAJ, Koob GF, Kieffer BL: Identification of genes regulated in the mouse extended amygdala by excessive ethanol drinking associated with dependence. *Addict Biol* 2011, 16:615–619.

157. Moonat S, Starkman BG, Sakharkar A, Pandey SC: Neuroscience of alcoholism: molecular and cellular mechanisms. *Cell Mol Life Sci* 2010, 67:73–88.

158. Clark PJ, Kohman RA, Miller DS, Bhattacharya TK, Brzezinska WJ, Rhodes JS: Genetic influences on exercise-induced adult hippocampal neurogenesis across 12 divergent mouse strains. *Genes Brain Behav* 2011, 10:345–353.

159. Leasure JL, Nixon K: Exercise Neuroprotection in a Rat Model of Binge Alcohol Consumption. *Alcohol Clin Exp Res* 2010, 34:404–414.

160. Oliff HS, Berchtold NC, Isackson P, Cotman CW: Exercise-induced regulation of brain-derived neurotrophic factor (BDNF) transcripts in the rat hippocampus. *Brain Res Mol Brain Res* 1998, 61:147–153.

161. Adlard PA, Perreau VM, Cotman CW: The exercise-induced expression of BDNF within the hippocampus varies across life-span. *Neurobiol Aging* 2005, 26:511–520.

162. Van Praag H: Neurogenesis and Exercise: Past and Future Directions. *Neuromolecular Med* 2008, 10:128–140.

163. Clark PJ, Kohman RA, Miller DS, Bhattacharya TK, Haferkamp EH, Rhodes JS: Adult hippocampal neurogenesis and c-Fos induction during escalation of voluntary wheel running in C57BL/6J mice. *Behavioural Brain Research* 2010, 213:246–252.

164. Kamens HM, Andersen J, Picciotto MR: Modulation of ethanol consumption by genetic and pharmacological manipulation of nicotinic acetylcholine receptors in mice. *Psychopharmacology (Berl)* 2010, 208:613–626.

165. Kamens HM, Burkhart-Kasch S, McKinnon CS, Li N, Reed C, Phillips TJ: Sensitivity to psychostimulants in mice bred for high and low stimulation to methamphetamine. *Genes Brain Behav* 2005, 4:110–125.

166. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001, 25:402–408.

167. Schmittgen TD, Livak KJ: Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 2008, 3:1101–1108.

168. Marks MJ, Pauly JR, Gross SD, Deneris ES, Hermans-Borgmeyer I, Heinemann SF, Collins AC: Nicotine binding and nicotinic receptor subunit RNA after chronic nicotine treatment. *J Neurosci* 1992, 12:2765–2784.

169. He D-YD, Ron DD: Glial cell line-derived neurotrophic factor reverses ethanol-mediated increases in tyrosine hydroxylase immunoreactivity via altering the activity of heat shock protein 90. *J Biol Chem* 2008, 283:12811–12818.

170. Imperato A, Di Chiara G: Preferential stimulation of dopamine release in the nucleus accumbens of freely moving rats by ethanol. *J Pharmacol Exp Ther* 1986, 239:219–228.

171. Everitt BJ, Robbins TW: Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci* 2005, 8:1481–1489.

172. Prado-Alcalá R, Wise RA: Brain stimulation reward and dopamine terminal fields. I. Caudate-putamen, nucleus accumbens and amygdala. *Brain Res* 1984, 297:265–273.

173. Wise RA: Intracranial self-stimulation: mapping against the lateral boundaries of the dopaminergic cells of the substantia nigra. *Brain Res* 1981, 213:190–194.

174. Wise RA: Roles for nigrostriatal--not just mesocorticolimbic--dopamine in reward and addiction. *Trends in Neurosciences* 2009, 32:517–524.

175. Vollstädt-Klein S, Wichert S, Rabinstein J, Bühler M, Klein O, Ende G, Hermann D, Mann K: Initial, habitual and compulsive alcohol use is characterized by a shift of cue processing from ventral to dorsal striatum. *Addiction* 2010, 105:1741–1749.

176. Blankenberg D, Gordon A, Kuster Von G, Coraor N, Taylor J, Nekrutenko A: Manipulation of FASTQ data with Galaxy. *…* 2010.

177. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010, 26:139–140.

178. Robinson MDM, Smyth GKG: Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 2007, 23:2881–2887.

179. Robinson MDM, Smyth GKG: Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* 2008, 9:321–332.

180. McCarthy DJ, Chen Y, Smyth GK: Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* 2012, 40:4288–4297.

181. R Core Team 2012: R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria ISBN: 3-900051-07-0 http://wwwR-projectorg/*.

182. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J: Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 2004, 5:R80–R80.

183. Greenwood BN, Foley TE, Day HEW, Campisi J, Hammack SH, Campeau S, Maier SF, Fleshner M: Freewheel running prevents learned helplessness/behavioral depression: role of dorsal raphe serotonergic neurons. *Journal of Neuroscience* 2003, 23:2889–2898.

184. Greenwood BN, Spence KG, Crevling DM, Clark PJ, Craig WC, Fleshner M: Exercise-induced stress resistance is independent of exercise controllability and the medial prefrontal cortex. *Eur J Neurosci* 2013, 37:469–478.

185. Greenwood BN, Fleshner M: Exercise, learned helplessness, and the stress-resistant brain. *Neuromolecular Med* 2008, 10:81–98.

186. Duman CH, Schlesinger L, Russell DS, Duman RS: Voluntary exercise produces antidepressant and anxiolytic behavioral effects in mice. *Brain Res* 2008, 1199:148–158.

187. Hunsberger JG, Newton SS, Bennett AH, Duman CH, Russell DS, Salton SR, Duman RS: Antidepressant actions of the exercise-regulated gene VGF. *Nat Med* 2007, 13:1476–1482.

188. Greenwood BN, Loughridge AB, Sadaoui N, Christianson JP, Fleshner M: The protective effects of voluntary exercise against the behavioral consequences of uncontrollable stress persist despite an increase in anxiety following forced cessation of exercise. *Behavioural Brain Research* 2012, 233:314–321.

189. Adlard PA, Cotman CW: Voluntary exercise protects against stress-induced decreases in brain-derived neurotrophic factor protein expression. *Neuroscience* 2004, 124:985–992.

190. Brene S, Bjørnebekk A, Aberg E, Mathé AA, Olson L, Werme M: Running is rewarding and antidepressive. *Physiology & Behavior* 2007, 92:136–140.

191. Kanarek RB, Marks-Kaufman R, D'Anci KE, Przypek J: Exercise attenuates oral intake of amphetamine in rats. *Pharmacol Biochem Behav* 1995, 51:725–729.

192. Cosgrove KP, Hunter RG, Carroll ME: Wheel-running attenuates intravenous cocaine self-administration in rats: sex differences. *Pharmacol Biochem Behav* 2002, 73:663–671.

193. Smith MA, Walker KL, Cole KT, Lang KC: The effects of aerobic exercise on cocaine self-administration in male and female rats. *Psychopharmacology (Berl)* 2011, 218:357–369.

194. Sanchez V, Moore CF, Brunzell DH, Lynch WJ: Effect of wheel-running during abstinence on subsequent nicotine-seeking in rats. *Psychopharmacology (Berl)* 2013.

195. Worst TJ, Tan JC, Robertson DJ, Freeman WM, Hyytia P, Kiianmaa K, Vrana KE: Transcriptome analysis of frontal cortex in alcohol-preferring and nonpreferring rats. *J Neurosci Res* 2005, 80:529–538.

196. Staley MJ, Naidoo D, Pridmore SA: Concentrations of transthyretin (prealbumin) and retinol-binding protein in alcoholics during alcohol withdrawal. *Clin Chem* 1984, 30:1887–1887.

Appendix

Ethanol Responsive

Gene Symbol	Gene name	locus	log Fold Change	p-value	FDR	QTL
Kcnj13	potassium inwardly-rectifying channel, subfamily J, member 13	chr1:87386363-87394729	-1.82	3.89E-14	6.14E-11	
Sostdc1	sclerostin domain containing 1	chr12:36314169-36318452	-1.66	1.39E-11	1.79E-08	
Col8a2	collagen, type VIII, alpha 2	chr4:126286793-126314330	-1.42	7.46E-09	6.62E-06	
Otx2	orthodenticle homolog 2 (Drosophila)	chr14:48657679-48667644	-1.19	2.98E-07	2.49E-04	
Col4a3	collagen, type IV, alpha 3	chr1:82586921-82722059	-1.31	8.67E-07	6.84E-04	
Sema3b	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B	chr9:107597674-107609229	-1.15	1.51E-06	1.13E-03	
Oca2	oculocutaneous albinism II	chr7:56239760-56536517	-1.24	1.06E-05	7.14E-03	
Ttc29	tetratricopeptide repeat domain 29	chr8:78213297-78394326	-1.18	1.64E-05	1.02E-02	
Wdr16	WD repeat domain 16	chr11:67924806-67965651	-1.04	1.73E-05	1.02E-02	
Slc2a12	solute carrier family 2 (facilitated glucose transporter), member 12	chr10:22645011-22704285	-0.97	1.85E-05	1.02E-02	
Dsp	desmoplakin	chr13:38151328-38198577	-1.17	1.86E-05	1.02E-02	
Ahdc1	AT hook, DNA binding motif, containing 1	chr4:133011260-133077863	0.90	2.80E-05	1.42E-02	
Ephx3	epoxide hydrolase 3	chr17:32183770-32189549	-1.48	4.33E-05	2.12E-02	
Acot4	acyl-CoA thioesterase 4	chr12:84038379-84044723	1.39	4.56E-05	2.16E-02	
Wdr86	WD repeat domain 86	chr5:24711738-24730727	-1.12	8.06E-05	3.58E-02	
Frem1	Fras1 related extracellular matrix protein 1	chr4:82897927-83052339	-0.90	1.03E-04	4.16E-02	
Icosl	icos ligand	chr10:78069368-78079525	0.91	1.20E-04	4.72E-02	
Gm10621	predicted gene 10621	chr9:108648840-108650953	1.20	1.29E-04	4.95E-02	

Exercise Responsive

Gene Symbol	Gene name	locus	log Fold Change	p-value	FDR	QTL
Mlf1	myeloid leukemia factor 1	chr3:67374097-67400003	-1.54	1.35E-08	1.06E-05	
Raph1	Ras association (RalGDS/AF-6) and pleckstrin homology domains 1	chr1:60490412-60567104	1.34	2.83E-08	2.01E-05	
Eif4ebp2	eukaryotic translation initiation factor 4E binding protein 2	chr10:61432497-61452669	1.40	5.16E-08	3.19E-05	
Nova2	neuro-oncological ventral antigen 2	chr7:18925888-18962057	1.17	6.58E-08	3.89E-05	
Kcnj13	potassium inwardly-rectifying channel, subfamily J, member 13	chr1:87386363-87394729	-1.25	9.68E-08	5.10E-05	
Lrtm1	leucine-rich repeats and transmembrane domains 1	chr14:29018208-29033642	1.18	1.10E-07	5.59E-05	
Proser1	proline and serine rich 1	chr3:53463666-53481755	1.16	1.60E-07	7.14E-05	
Sostdc1	sclerostin domain containing 1	chr12:36314169-36318452	-1.27	1.61E-07	7.14E-05	
Otx2	orthodenticle homolog 2 (Drosophila)	chr14:48657679-48667644	-1.23	2.06E-07	8.87E-05	
Mll2	myeloid/lymphoid or mixed-lineage leukemia 2	chr15:98831669-98871183	1.08	3.36E-07	1.36E-04	
Krt18	keratin 18	chr15:102028216-102032026	-1.32	8.58E-07	3.10E-04	
Tcf7l1	transcription factor 7 like 1 (T cell specific, HMG box)	chr6:72626378-72789254	1.26	8.83E-07	3.10E-04	
Gltscr1	glioma tumor suppressor candidate region gene 1	chr7:15971262-15999495	1.12	1.05E-06	3.36E-04	
Oca2	oculocutaneous albinism II	chr7:56239760-56536517	-1.45	1.06E-06	3.36E-04	
Ccdc153	coiled-coil domain containing 153	chr9:44240677-44247306	-1.13	1.06E-06	3.36E-04	*Etp5*
Bcl9l	B cell CLL/lymphoma 9-like	chr9:44499136-44510388	1.06	1.13E-06	3.43E-04	*Etp5*
Lrrc34	leucine rich repeat containing 34	chr3:30624267-30647869	-1.52	1.15E-06	3.43E-04	
Kcnj6	potassium inwardly-rectifying channel, subfamily J, member 6	chr16:94749266-94997696	1.24	1.16E-06	3.43E-04	
Nfic	nuclear factor I/C	chr10:81396186-81431005	1.03	1.51E-06	4.38E-04	
Prr12	proline rich 12	chr7:45027707-45052881	1.03	1.71E-06	4.81E-04	
Ids	iduronate 2-sulfatase	chrX:70343069-70365084	1.01	1.75E-06	4.81E-04	
Tsnaxip1	translin-associated factor X (Tsnax) interacting protein 1	chr8:105827744-105844676	-1.49	1.76E-06	4.81E-04	
Ybx2	Y box protein 2	chr11:69935796-69941605	1.55	1.79E-06	4.81E-04	
Mbd6	methyl-CpG binding domain protein 6	chr10:127281956-127289018	1.02	2.07E-06	5.45E-04	
Fam160b2	family with sequence similarity 160, member B2	chr14:70583296-70599835	1.00	2.27E-06	5.75E-04	
Zfp871	zinc finger protein 871	chr17:32771236-32788287	1.00	2.72E-06	6.66E-04	
Trim25	tripartite motif-containing 25	chr11:88999376-89020293	1.08	2.80E-06	6.74E-04	
Clic6	chloride intracellular channel 6	chr16:92485736-92541243	-1.01	2.93E-06	6.87E-04	
Cbl	Casitas B-lineage lymphoma	chr9:44149262-44234046	1.02	3.13E-06	7.11E-04	*Etp5*

Exercise cont. Gene Symbol	Gene name	locus	log Fold Change	p-value	FDR	QTL
Scrt1	scratch homolog 1, zinc finger protein (Drosophila)	chr15:76516203-76522129	1.00	3.15E-06	7.11E-04	
Sap30	sin3 associated polypeptide	chr8:57482707-57487860	-1.09	3.21E-06	7.11E-04	
Runx3	runt related transcription factor 3	chr4:135120652-135177990	1.20	3.68E-06	8.04E-04	
Ccer1	coiled coil glutamate rich protein 1	chr10:97693059-97694926	1.68	3.95E-06	8.50E-04	
Dcc	deleted in colorectal carcinoma	chr18:71258738-72351069	0.99	4.39E-06	9.31E-04	
Lbp	lipopolysaccharide binding protein	chr2:158306493-158332852	-1.06	6.90E-06	1.42E-03	
Xkrx	X Kell blood group precursor related X linked	chrX:134149043-134162076	1.08	7.90E-06	1.60E-03	
Slc2a12	solute carrier family 2 (facilitated glucose transporter), member 12	chr10:22645011-22704285	-1.02	8.56E-06	1.69E-03	
Col9a3	collagen, type IX, alpha 3	chr2:180597790-180622189	-0.97	9.59E-06	1.87E-03	
Ikzf4	IKAROS family zinc finger 4	chr10:128630843-128645991	0.96	1.02E-05	1.95E-03	
Scrt2	scratch homolog 2, zinc finger protein (Drosophila)	chr2:152081529-152095802	0.96	1.03E-05	1.95E-03	*Etp1*
Spen	SPEN homolog, transcriptional regulator (Drosophila)	chr4:141467890-141538597	0.94	1.25E-05	2.31E-03	
Sema3b	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B	chr9:107597674-107609229	-1.06	1.25E-05	2.31E-03	
Uqcrq	ubiquinol-cytochrome c reductase, complex III subunit VII	chr11:53427922-53430831	-0.92	1.37E-05	2.42E-03	
Kdm6b	KDM1 lysine (K)-specific demethylase 6B	chr11:69398508-69413675	0.96	1.38E-05	2.42E-03	
Col8a2	collagen, type VIII, alpha 2	chr4:126286793-126314330	-1.06	1.44E-05	2.50E-03	
Frem1	Fras1 related extracellular matrix protein 1	chr4:82897927-83052339	-1.02	1.46E-05	2.50E-03	
Srcap	Snf2-related CREBBP activator protein	chr7:127511983-127566940	0.93	1.50E-05	2.53E-03	
Ccdc146	coiled-coil domain containing 146	chr5:21292961-21424677	-1.05	1.61E-05	2.69E-03	
Fam57b	family with sequence similarity 57, member B	chr7:126816885-126830219	0.96	1.72E-05	2.81E-03	
Krt20	keratin 20	chr11:99428403-99438150	1.13	1.90E-05	3.03E-03	
Gm9796	predicted gene 9796	chr11:95696898-95699143	1.34	1.94E-05	3.03E-03	
Fhad1	forkhead-associated (FHA) phosphopeptide binding domain 1	chr4:141890438-142015082	-0.94	1.94E-05	3.03E-03	
Gm10287	predicted gene 10287	chr3:149221693-149225745	1.31	2.02E-05	3.11E-03	
Myl4	myosin, light polypeptide 4	chr11:104550663-104595753	-0.91	2.30E-05	3.50E-03	
Arhgap33	Rho GTPase activating protein 33	chr7:30522226-30534180	0.90	2.32E-05	3.50E-03	
Smad6	SMAD family member 6	chr9:63953076-64022059	-1.13	2.39E-05	3.50E-03	*Etp5*
Shisa6	shisa homolog 6 (Xenopus laevis)	chr11:66211725-66525795	0.96	2.39E-05	3.50E-03	
Arc	activity regulated cytoskeletal-associated protein	chr15:74669083-74672570	-0.89	2.43E-05	3.53E-03	

Exercise cont. Gene Symbol	Gene name	locus	log Fold Change	p-value	FDR	QTL
Slc9b1	solute carrier family 9, subfamily B (NHA1, cation proton antiporter 1), member 1	chr3:135348029-135397827	1.41	2.54E-05	3.64E-03	
Muc19	mucin 19	chr15:91838326-91936388	1.26	2.57E-05	3.64E-03	
1700003M02Rik	RIKEN cDNA 1700003M02 gene	chr4:34688559-34730206	-1.02	2.59E-05	3.64E-03	
Col4a3	collagen, type IV, alpha 3	chr1:82586921-82722059	-1.13	2.67E-05	3.72E-03	
Syn1	synapsin I	chrX:20860511-20921004	0.88	2.73E-05	3.77E-03	
Ppargc1b	peroxisome proliferative activated receptor, gamma, coactivator 1 beta	chr18:61298136-61400431	0.93	2.83E-05	3.87E-03	
Erich2	glutamate rich 2	chr2:70508819-70540884	-1.15	2.86E-05	3.87E-03	*Etp1*
Dot1l	DOT1-like, histone H3 methyltransferase (S. cerevisiae)	chr10:80755206-80795461	0.89	3.02E-05	4.05E-03	
St8sia2	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2	chr7:73939119-74013682	0.89	3.15E-05	4.19E-03	
Tjp3	tight junction protein 3	chr10:81273207-81291267	-1.81	3.27E-05	4.30E-03	
Hist1h4d	histone cluster 1, H4d	chr13:23581598-23581990	0.92	3.40E-05	4.44E-03	
Cacng3	calcium channel, voltage-dependent, gamma subunit 3	chr7:122671744-122769391	0.88	3.45E-05	4.46E-03	
Capsl	calcyphosine-like	chr15:9436028-9466035	-1.12	3.62E-05	4.63E-03	
BC100451	cDNA sequence BC100451	chr11:118332360-118342500	1.35	3.73E-05	4.73E-03	
Gatad2b	GATA zinc finger domain containing 2B	chr3:90341654-90358120	0.89	3.84E-05	4.83E-03	
Zmiz1	zinc finger, MIZ-type containing 1	chr14:25459185-25666743	0.87	3.99E-05	4.98E-03	
Stx1b	syntaxin 1B	chr7:127803900-127824549	0.86	4.07E-05	5.03E-03	
Srrm2	serine/arginine repetitive matrix 2	chr17:23803187-23824739	0.86	4.63E-05	5.64E-03	
Gm9930	predicted gene 9930	chr10:9532531-9535681	1.01	4.65E-05	5.64E-03	
Mospd3	motile sperm domain containing 3	chr5:137596645-137601058	1.02	4.70E-05	5.66E-03	
Egr2	early growth response 2	chr10:67535475-67542188	-0.89	5.07E-05	6.01E-03	
Lnpep	leucyl/cystinyl aminopeptidase	chr17:17527723-17624489	0.87	5.08E-05	6.01E-03	
Top2a	topoisomerase (DNA) II alpha	chr11:98992943-99024189	-0.89	5.22E-05	6.13E-03	
Nfat5	nuclear factor of activated T cells 5	chr8:107293470-107379517	0.85	5.64E-05	6.57E-03	
Ankrd52	ankyrin repeat domain 52	chr10:128377115-128408704	0.86	5.74E-05	6.63E-03	
Psma8	proteasome (prosome, macropain) subunit, alpha type, 8	chr18:14706151-14762299	1.58	6.35E-05	7.27E-03	
Ankrd9	ankyrin repeat domain 9	chr12:110975353-110979040	-0.98	7.51E-05	8.50E-03	
Pigr	polymeric immunoglobulin receptor	chr1:130826684-130852249	1.07	7.54E-05	8.50E-03	
Il21r	interleukin 21 receptor	chr7:125603429-125633570	1.13	7.68E-05	8.53E-03	
Srgap3	SLIT-ROBO Rho GTPase activating protein 3	chr6:112717971-112947266	0.83	7.69E-05	8.53E-03	

Exercise cont. Gene Symbol	Gene name	locus	log Fold Change	p-value	FDR	QTL
Ctxn1	cortexin 1	chr8:4257648-4259274	0.83	8.74E-05	9.55E-03	
Usp17la	ubiquitin specific peptidase 17-like A	chr7:104857009-104862667	1.30	9.12E-05	9.89E-03	
Cdk5r2	cyclin-dependent kinase 5, regulatory subunit 2 (p39)	chr1:74854934-74857431	0.82	9.42E-05	1.01E-02	
Nav2	neuron navigator 2	chr7:49246189-49610087	0.83	9.63E-05	1.03E-02	
Gpr63	G protein-coupled receptor 63	chr4:24966407-25009233	0.98	9.99E-05	1.06E-02	
Spred2	sprouty-related, EVH1 domain containing 2	chr11:19924375-20024026	0.83	1.03E-04	1.08E-02	
Myo1h	myosin 1H	chr5:114314941-114364576	1.29	1.03E-04	1.08E-02	
Tnfrsf22	tumor necrosis factor receptor superfamily, member 22	chr7:143634808-143649661	0.82	1.05E-04	1.09E-02	
Stx19	syntaxin 19	chr16:62814676-62824346	1.14	1.06E-04	1.09E-02	
A430110L20Rik	RIKEN cDNA A430110L20 gene	chr1:181226076-181228490	0.88	1.16E-04	1.19E-02	
Rorc	RAR-related orphan receptor gamma	chr3:94372794-94398276	-0.99	1.17E-04	1.19E-02	
Wdr96	WD repeat domain 96	chr19:47737561-47919287	-0.88	1.23E-04	1.24E-02	
Wnt2b	wingless related MMTV integration site 2b	chr3:104944805-104961709	0.92	1.25E-04	1.25E-02	
Nek5	NIMA (never in mitosis gene a)-related expressed kinase 5	chr8:22073616-22125053	-1.05	1.27E-04	1.26E-02	
Gucy1a2	guanylate cyclase 1, soluble, alpha 2	chr9:3532354-3897342	0.82	1.32E-04	1.30E-02	
Nxpe2	neurexophilin and PC-esterase domain family, member 2	chr9:48318006-48353454	1.01	1.36E-04	1.32E-02	*Etp5*
Ak7	adenylate kinase 7	chr12:105705982-105782447	-0.86	1.36E-04	1.32E-02	
Vamp2	vesicle-associated membrane protein 2	chr11:69088490-69092384	0.80	1.37E-04	1.32E-02	
Gm14308	predicted gene 14308	chr2:176613364-176636344	1.00	1.37E-04	1.32E-02	
Zfp385a	zinc finger protein 385A	chr15:103313895-103340086	0.83	1.40E-04	1.33E-02	
Acr	acrosin prepropeptide	chr15:89568326-89574585	1.00	1.41E-04	1.33E-02	
Lamtor4	late endosomal/lysosomal adaptor, MAPK and MTOR activator 4	chr5:138255608-138259398	-0.82	1.41E-04	1.33E-02	
Ttc29	tetratricopeptide repeat domain 29	chr8:78213297-78394326	-1.06	1.47E-04	1.38E-02	
Mylk4	myosin light chain kinase family, member 4	chr13:32704680-32783954	1.07	1.54E-04	1.43E-02	
Rhox8	reproductive homeobox 8	chrX:37874780-37878944	1.22	1.55E-04	1.43E-02	
Rreb1	ras responsive element binding protein 1	chr13:37778400-37952002	0.80	1.56E-04	1.43E-02	
Sipa1l1	signal-induced proliferation-associated 1 like 1	chr12:82170016-82451782	0.79	1.67E-04	1.51E-02	
Tcp11l1	t-complex 11 like 1	chr2:104657288-104712169	0.80	1.69E-04	1.52E-02	*Etp1*
Hist1h2bg	histone cluster 1, H2bg	chr13:23571408-23571884	0.84	1.79E-04	1.60E-02	
Gm10576	predicted gene 10576	chr4:101054558-101055262	0.97	1.82E-04	1.62E-02	
Gm1078	predicted gene 1078	chr7:4965260-4971168	1.00	1.86E-04	1.64E-02	

Exercise cont. Gene Symbol	Gene name	locus	log Fold Change	p-value	FDR	QTL
Apex2	apurinic/apyrimidinic endonuclease 2	chrX:150519519-150589868	0.81	1.91E-04	1.68E-02	
Wdr16	WD repeat domain 16	chr11:67924806-67965651	-0.91	1.94E-04	1.69E-02	
Rgs22	regulator of G-protein signalling 22	chr15:36009479-36140400	-0.91	1.96E-04	1.70E-02	
Psg23	pregnancy-specific glycoprotein 23	chr7:18606343-18616501	0.94	2.02E-04	1.74E-02	
Mfrp	membrane-type frizzled-related protein	chr9:44101738-44109187	-0.84	2.04E-04	1.74E-02	*Etp5*
Gpr18	G protein-coupled receptor 18	chr14:121911435-121915774	1.14	2.12E-04	1.80E-02	
Gm6104	predicted gene 6104	chr1:4879208-4880663	1.31	2.24E-04	1.89E-02	
A130050O07Rik	RIKEN cDNA A130050O07 gene	chr1:137928170-137930273	0.98	2.28E-04	1.91E-02	
Tmc5	transmembrane channel-like gene family 5	chr7:118597297-118675086	0.84	2.28E-04	1.91E-02	
Muc15	mucin 15	chr2:110721340-110739527	0.83	2.31E-04	1.91E-02	*Etp1*
Cox5b	cytochrome c oxidase subunit Vb	chr1:36691487-36693385	-0.77	2.31E-04	1.91E-02	
Lrp2	low density lipoprotein receptor-related protein 2	chr2:69424340-69586065	-0.96	2.33E-04	1.91E-02	*Etp1*
Setd1b	SET domain containing 1B	chr5:123142193-123167435	0.81	2.34E-04	1.91E-02	
Ahdc1	AT hook, DNA binding motif, containing 1	chr4:133011260-133077863	0.79	2.36E-04	1.91E-02	
A930011G23Rik	RIKEN cDNA A930011G23 gene	chr5:99297244-99729065	0.82	2.42E-04	1.95E-02	
Pcsk1n	proprotein convertase subtilisin/kexin type 1 inhibitor	chrX:7919822-7924410	0.77	2.58E-04	2.07E-02	
Fos	FBJ osteosarcoma oncogene	chr12:85473890-85477273	-0.81	2.65E-04	2.11E-02	
Mid1	midline 1	chrX:169685199-169990798	0.77	2.87E-04	2.28E-02	
Ttk	Ttk protein kinase	chr9:83834689-83872390	-1.16	2.89E-04	2.28E-02	*Etp5*
Gpx8	glutathione peroxidase 8 (putative)	chr13:113042763-113046388	-0.83	2.93E-04	2.30E-02	
Darc	Duffy blood group, chemokine receptor	chr1:173331886-173333503	0.79	3.03E-04	2.37E-02	
Frmpd3	FERM and PDZ domain containing 3	chrX:140367494-140394540	1.00	3.21E-04	2.48E-02	
Sulf1	sulfatase 1	chr1:12692430-12860371	-0.77	3.22E-04	2.48E-02	
Fam78b	family with sequence similarity 78, member B	chr1:167001432-167091009	0.76	3.27E-04	2.51E-02	
Myrf	myelin regulatory factor	chr19:10208272-10240748	0.76	3.34E-04	2.55E-02	
Pdlim7	PDZ and LIM domain 7	chr13:55495795-55513676	0.77	3.38E-04	2.56E-02	
Naip5	NLR family, apoptosis inhibitory protein 5	chr13:100211739-100246323	0.91	3.42E-04	2.58E-02	
Hdx	highly divergent homeobox	chrX:111569931-111697079	0.79	3.43E-04	2.58E-02	
Bcl9	B cell CLL/lymphoma 9	chr3:97203662-97228846	0.77	3.52E-04	2.63E-02	
A830073O21Rik	RIKEN cDNA A830073O21 gene	chr7:73738893-73740917	0.79	3.57E-04	2.65E-02	
Calml4	calmodulin-like 4	chr9:62858104-62875918	-0.87	3.63E-04	2.67E-02	*Etp5*
Rnf219	ring finger protein 219	chr14:104477534-104522666	0.77	3.66E-04	2.68E-02	
Wdr93	WD repeat domain 93	chr7:79743163-79785950	-1.09	3.71E-04	2.70E-02	
Lrrc29	leucine rich repeat containing 29	chr8:105312341-105326276	1.05	3.75E-04	2.71E-02	

Exercise cont. Gene Symbol	Gene name	locus	log Fold Change	p-value	FDR	QTL
Nrg3	neuregulin 3	chr14:38368952-39473088	0.76	3.76E-04	2.71E-02	
Pdk2	pyruvate dehydrogenase kinase, isoenzyme 2	chr11:95026258-95041354	0.75	3.80E-04	2.72E-02	
Atf7	activating transcription factor 7	chr15:102536643-102625421	0.78	4.06E-04	2.89E-02	
1700001L05Rik	RIKEN cDNA 1700001L05 gene	chr15:83357526-83367282	1.04	4.07E-04	2.89E-02	
Cd33	CD33 antigen	chr7:43527456-43533171	0.80	4.19E-04	2.96E-02	
Fam65a	family with sequence similarity 65, member A	chr8:105605229-105622194	0.75	4.25E-04	2.98E-02	
Nfix	nuclear factor I/X	chr8:84699876-84800344	0.75	4.25E-04	2.98E-02	
Zkscan16	zinc finger with KRAB and SCAN domains 16	chr4:58943628-58958355	0.75	4.45E-04	3.10E-02	
Scin	scinderin	chr12:40059769-40134228	-1.29	4.47E-04	3.10E-02	
Ccdc33	coiled-coil domain containing 33	chr9:58028677-58118823	-0.86	4.51E-04	3.11E-02	*Etp5*
Zdhhc23	zinc finger, DHHC domain containing 23	chr16:43969146-43979050	0.87	4.58E-04	3.14E-02	
Gm10033	predicted gene 10033	chr8:69372145-69373383	-0.82	4.64E-04	3.17E-02	
Zfp579	zinc finger protein 579	chr7:4983483-4996158	0.78	4.68E-04	3.18E-02	
4921522P10Rik	RIKEN cDNA 4921522P10 gene	chr8:8661801-8664728	0.98	4.91E-04	3.32E-02	
Sp8	trans-acting transcription factor 8	chr12:118846329-118852578	-0.92	4.97E-04	3.34E-02	
Dsc3	desmocollin 3	chr18:19960930-20002097	-0.95	5.05E-04	3.38E-02	
Myadm	myeloid-associated differentiation marker	chr7:3289038-3299345	0.74	5.18E-04	3.45E-02	
4930555F03Rik	RIKEN cDNA 4930555F03 gene	chr8:49370886-49521095	1.00	5.31E-04	3.53E-02	
Zfp609	zinc finger protein 609	chr9:65692391-65827564	0.73	5.40E-04	3.57E-02	*Etp5*
Dsp	desmoplakin	chr13:38151328-38198577	-0.96	5.57E-04	3.66E-02	
Zmat3	zinc finger matrin type 3	chr3:32334798-32365678	0.73	5.74E-04	3.76E-02	
Pfdn4	prefoldin 4	chr2:170496428-170519123	-0.73	5.81E-04	3.78E-02	
Cd24a	CD24a antigen	chr10:43579169-43584262	-0.73	5.87E-04	3.81E-02	
Oprm1	opioid receptor, mu 1	chr10:6758506-7038198	0.83	5.99E-04	3.87E-02	
Fam183b	family with sequence similarity 183, member B	chr11:58792797-58801960	-0.90	6.17E-04	3.96E-02	
Lcor	ligand dependent nuclear receptor corepressor	chr19:41549639-41559781	0.73	6.25E-04	4.00E-02	
Tet3	tet methylcytosine dioxygenase 3	chr6:83362373-83441678	0.73	6.36E-04	4.03E-02	
H3f3a	H3 histone, family 3A	chr1:180800832-180813943	0.72	6.41E-04	4.05E-02	
Tcp10b	t-complex protein 10b	chr17:13061104-13082481	1.11	6.95E-04	4.37E-02	
Zfp703	zinc finger protein 703	chr8:26977336-26981461	0.82	7.12E-04	4.45E-02	
Zmiz2	zinc finger, MIZ-type containing 2	chr11:6389074-6406158	0.71	7.17E-04	4.47E-02	
Ankrd61	ankyrin repeat domain 61	chr5:143890741-143897685	0.88	7.21E-04	4.47E-02	
Tc2n	tandem C2 domains, nuclear	chr12:101645443-101718523	-0.99	7.37E-04	4.52E-02	
Dnahc6	dynein, axonemal, heavy chain 6	chr6:73017609-73221651	-0.76	7.38E-04	4.52E-02	
Armc4	armadillo repeat containing 4	chr18:7088233-7297901	-0.99	7.39E-04	4.52E-02	

Exercise cont. Gene Symbol	Gene name	locus	log Fold Change	p-value	FDR	QTL
Uhrf1	ubiquitin-like, containing PHD and RING finger domains, 1	chr17:56303321-56323486	-0.89	7.51E-04	4.56E-02	
Wfikkn1	WAP, FS, Ig, KU, and NTR-containing protein 1	chr17:25877630-25880305	1.22	7.53E-04	4.56E-02	
Pdlim3	PDZ and LIM domain 3	chr8:45885485-45919546	-0.83	7.55E-04	4.56E-02	
Dnali1	dynein, axonemal, light intermediate polypeptide 1	chr4:125055338-125065703	-0.82	7.61E-04	4.58E-02	
Daw1	dynein assembly factor with WDR repeat domains 1	chr1:83159752-83210574	-0.99	7.68E-04	4.59E-02	
Clrn1	clarin 1	chr3:58844028-58885340	1.40	7.69E-04	4.59E-02	
Shisa7	shisa homolog 7 (Xenopus laevis)	chr7:4825552-4836723	0.71	7.76E-04	4.61E-02	
Diras1	DIRAS family, GTP-binding RAS-like 1	chr10:81019589-81025662	0.70	8.07E-04	4.78E-02	
Wnk4	WNK lysine deficient protein kinase 4	chr11:101260567-101277409	0.74	8.26E-04	4.87E-02	
Rpp25l	ribonuclease P/MRP 25 subunit-like	chr4:41712033-41713534	-0.75	8.40E-04	4.93E-02	
Cd200r3	CD200 receptor 3	chr16:44943678-44981380	1.07	8.45E-04	4.93E-02	
Clip3	CAP-GLY domain containing linker protein 3	chr7:30291672-30308367	0.70	8.47E-04	4.93E-02	
Gm4787	predicted gene 4787	chr12:81377136-81379382	0.83	8.50E-04	4.93E-02	
Sec22c	SEC22 vesicle trafficking protein homolog C (S. cerevisiae)	chr9:121683022-121705490	0.74	8.56E-04	4.94E-02	
Fam216b	family with sequence similarity 216, member B	chr14:78081021-78089007	-0.79	8.63E-04	4.96E-02	

Interaction Gene Symbol	Gene name	locus	p-value	FDR	QTL
Cdr1	cerebellar degeneration related antigen 1	chrX:61183246-61185558	2.76E-36	3.92E-32	
Prkcd	protein kinase C, delta	chr14:30595358-30626210	1.27E-19	9.05E-16	
Tnnt1	troponin T1, skeletal, slow	chr7:4504570-4516382	3.16E-14	1.50E-10	
Lhx9	LIM homeobox protein 9	chr1:138825186-138848576	2.57E-11	9.14E-08	
Kcne2	potassium voltage-gated channel, Isk-related subfamily, gene 2	chr16:92292389-92298129	1.05E-10	2.98E-07	
Ramp3	receptor (calcitonin) activity modifying protein 3	chr11:6658521-6677475	1.44E-10	3.41E-07	
Vipr2	vasoactive intestinal peptide receptor 2	chr12:116077726-116146261	4.09E-10	8.30E-07	
Polr2a	polymerase (RNA) II (DNA directed) polypeptide A	chr11:69733997-69758637	1.26E-09	2.24E-06	
Slc13a4	solute carrier family 13 (sodium/sulfate symporters), member 4	chr6:35267957-35308131	7.79E-09	1.23E-05	
Slc17a6	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 6	chr7:51621830-51671125	1.22E-08	1.73E-05	
F5	coagulation factor V	chr1:164151838-164220277	2.02E-08	2.61E-05	
Folr1	folate receptor 1 (adult)	chr7:101858331-101870788	5.29E-08	6.20E-05	
Slc4a5	solute carrier family 4, sodium bicarbonate cotransporter, member 5	chr6:83219828-83304945	5.67E-08	6.20E-05	
Adamts19	a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 19	chr18:58836764-59053678	8.89E-08	9.02E-05	
Prlr	prolactin receptor	chr15:10177238-10349180	1.02E-07	9.70E-05	
Ly86	lymphocyte antigen 86	chr13:37345345-37419036	1.29E-07	1.14E-04	
Rhog	ras homolog gene family, member G	chr7:102239123-102250123	2.90E-07	2.42E-04	
Kl	klotho	chr5:150952607-150993809	3.51E-07	2.77E-04	
Nlrp1a	NLR family, pyrin domain containing 1A	chr11:71092236-71144704	4.35E-07	3.25E-04	
Epn3	epsin 3	chr11:94489599-94499974	4.92E-07	3.49E-04	
Shank1	SH3/ankyrin domain gene 1	chr7:44310253-44358351	5.32E-07	3.60E-04	
Rgs16	regulator of G-protein signaling 16	chr1:153740349-153745468	8.99E-07	5.81E-04	
Tmem72	transmembrane protein 72	chr6:116692630-116716913	2.16E-06	1.34E-03	
Col8a1	collagen, type VIII, alpha 1	chr16:57624258-57754737	2.35E-06	1.39E-03	
Ttr	transthyretin	chr18:20665250-20674324	3.12E-06	1.77E-03	
Abca4	ATP-binding cassette, sub-family A (ABC1), member 4	chr3:122044443-122180061	3.62E-06	1.98E-03	
Atf7	activating transcription factor 7	chr15:102536643-102625421	5.09E-06	2.68E-03	
Lrrc17	leucine rich repeat containing 17	chr5:21543527-21575900	5.49E-06	2.79E-03	
Fam163b	family with sequence similarity 163, member B	chr2:27110380-27142491	6.79E-06	3.33E-03	*Etp1*
Kcnk9	potassium channel, subfamily K, member 9	chr15:72512119-72546279	1.88E-05	8.88E-03	

Interaction cont.

Gene Symbol	Gene name	locus	p-value	FDR	QTL
Adam33	a disintegrin and metallopeptidase domain 33	chr2:131050591-131063814	3.48E-05	1.58E-02	*Etp1*
Zfhx2	zinc finger homeobox 2	chr14:55060262-55092324	3.57E-05	1.58E-02	
Zmynd15	zinc finger, MYND-type containing 15	chr11:70459433-70466202	3.67E-05	1.58E-02	
Gm10800	predicted gene 10800	chr2:98666547-98667301	4.15E-05	1.74E-02	*Etp1*
Tmem86b	transmembrane protein 86B	chr7:4628042-4630482	4.91E-05	1.99E-02	
Kcng3	potassium voltage-gated channel, subfamily G, member 3	chr17:83585957-83631895	5.43E-05	2.14E-02	
Ksr2	kinase suppressor of ras 2	chr5:117414000-117555942	5.92E-05	2.27E-02	
Zfp580	zinc finger protein 580	chr7:5051532-5053722	6.24E-05	2.33E-02	
Pi15	peptidase inhibitor 15	chr1:17601901-17630938	7.19E-05	2.62E-02	
Ccdc135	coiled-coil domain containing 135	chr8:95055103-95078141	7.59E-05	2.70E-02	
Apbb1ip	amyloid beta (A4) precursor protein-binding, family B, member 1 interacting protein	chr2:22774094-22875653	8.63E-05	2.95E-02	*Etp1*
4930566N20Rik	RIKEN cDNA 4930566N20 gene	chr3:157207662-157208860	8.73E-05	2.95E-02	
Hnf1a	HNF1 homeobox A	chr5:114948361-114971067	8.98E-05	2.97E-02	
Frem3	Fras1 related extracellular matrix protein 3	chr8:80611080-80695356	9.37E-05	3.02E-02	
4930481A15Rik	RIKEN cDNA 4930481A15 gene	chr19:5406740-5422847	1.02E-04	3.18E-02	
Igfbp2	insulin-like growth factor binding protein 2	chr1:72824503-72852474	1.03E-04	3.18E-02	
Pde6h	phosphodiesterase 6H, cGMP-specific, cone, gamma	chr6:136954523-136968865	1.08E-04	3.26E-02	
Tmprss11a	transmembrane protease, serine 11a	chr5:86410410-86468990	1.12E-04	3.30E-02	
A2m	alpha-2-macroglobulin	chr6:121636173-121679237	1.16E-04	3.36E-02	
Tox2	TOX high mobility group box family member 2	chr2:163203125-163324170	1.24E-04	3.52E-02	
Gm10855	predicted gene 10855	chr2:6932541-6935081	1.46E-04	4.06E-02	
Gja6	gap junction protein, alpha 6	chrX:160902116-160907052	1.54E-04	4.20E-02	
Ucp2	uncoupling protein 2 (mitochondrial, proton carrier)	chr7:100493337-100502020	1.65E-04	4.43E-02	