Published in final edited form as:

Alcohol Clin Exp Res. 2018 August ; 42(8): 1454–1465. doi:10.1111/acer.13777.

Gender Specific Effects of Selection for Drinking in the Dark on the Network Roles of Coding and Non-Coding RNAs

Ovidiu Dan Iancu¹, Alex Colville¹, Beth Wilmot¹, Robert Searles², Priscila Darakjian¹, Christina Zheng^{3,4}, Shannon McWeeney³, Sunita Kawane³, John C. Crabbe^{1,5}, Pamela Metten^{1,5}, Denesa Oberbeck¹, Robert Hitzemann¹

¹Department of Behavioral Neuroscience, Oregon Health & Science University, Portland, Oregon.

²Integrated Genomics Laboratory, Oregon Health & Science University, Portland, Oregon.

³Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon.

⁴Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon.

⁵VA Portland Health Care System, Portland, Oregon.

Abstract

Background: Transcriptional differences between Heterogeneous stock (HS/NPT) mice and High Drinking in the Dark (HDID) selected mouse lines have previously been described based on microarray technology coupled with network based analysis (Iancu et al., 2013b). The network changes were reproducible in two independent selections and largely confined to two distinct network modules; in contrast differential expression appeared more specific to each selected line. The current study extends these results by utilizing RNA-Seq technology, allowing evaluation of the relationship between genetic risk and transcription of non-coding RNA; we additionally evaluate sex-specific transcriptional effects of selection.

Methods: Naïve mice (N = 24/group and sex) were utilized for gene expression analysis in the ventral striatum; the transcriptome was sequenced with the Illumina HiSeq platform. Differential gene expression and the weighted gene coexpression network analysis (WGCNA) were implemented largely as described elsewhere (Colville et al., 2017), resulting in the identification of genes that change expression level or (co)variance structure.

Results: Across both sexes we detect selection effects on the extracellular matrix (ECM) and synaptic signaling, although the identity of individual genes varies. A majority of non-coding RNAs cluster in a single module of relatively low density in both the male and female network. The most strongly differentially expressed transcript in both sexes was *Gm22513*, a snRNA with unknown function. Associated with selection we also found a number of network hubs that change edge strength and connectivity. At the individual gene level there are many sex specific effects; however at the annotation level results are more concordant.

Keywords

Alcohol; drinking in the dark; binge; mouse; differential network analysis; non-coding RNA; selective breeding; sex specific effects

Introduction

The drinking-in-the-dark (DID) behavioral test was developed as a murine model of binge ethanol consumption (Rhodes et al., 2005). The phenotype is assessed by replacing the water bottle for singly-housed animals with a solution of 20% ethanol, starting 3–4 hours into their circadian dark phase; animals are allowed to drink the ethanol solution for the next 2–4 hours. The advantage of the DID model over other rodent models of ethanol drinking such as preference consumption (e.g. 24 hour, 2-bottle choice) is that some mouse genotypes will consume a sufficient quantity of ethanol to reach an intoxicating blood ethanol concentration during the DID test (> 80 mg%) (Rhodes et al., 2007). Our approach for investigating the neurobiology associated with DID has been to focus on the mouse lines selectively bred for their high blood alcohol levels, the high DID (HDID) selected lines (Barkley-Levenson and Crabbe, 2012a; 2014), to determine how the selected lines' brain transcriptomes changed from the genetically heterogeneous stock founders (HS/NPT) (Iancu et al., 2013b). In studies performed earlier, microarrays were used to collect gene expression data from specific brain regions in naïve mice, and a network-based approach (Zhang and Horvath, 2005) was used for data analysis. Several network co-expression modules were identified; two (denoted as black and magenta) were highly enriched in neuronal genes. In both modules, neuronal GO annotations such as glutamate secretion and trans-synaptic signaling were markedly disrupted by selection. Key hub genes affected included *Camkk2*, *Dgkz*, *Gabrg1*, *Gla2* and *Grik1*. Overall, the data pointed to synaptic remodeling as being associated with excessive binge ethanol consumption (Iancu et al, 2013b). These results are conceptually similar to those we recently obtained from nucleus accumbens tissue in HS-CC mice by comparing those mice with high and low preference consumption (Colville et al., 2017), although the specific genes in the co-expression modules differ between the two studies.

The purpose of the current study was to use a RNA-Seq based approach to extend our transcriptome analyses. RNA-Seq has a number of advantages over microarray-based approaches for interrogating the transcriptome (see e.g., (Hitzemann et al., 2013; 2014). These include greater sensitivity for detecting gene expression (Bottomly et al. 2011), a greater dynamic range for detecting differential gene expression, a variance structure that is better suited to data clustering and network analysis (Iancu et al. 2012) and the ability to globally assess changes in the expression of non-coding RNAs (ncRNAs).

The roles of ncRNAs are predominantly regulatory; playing complex and crucial roles in cell function, developmental regulation, and evolutionary expansion as well as disease initiation and progression (Guennewig and Cooper, 2014). ncRNAs achieve this by regulating gene expression at a number of levels including epigenetic modification, enhancer function, alternative RNA splicing, and translation (Mercer et al., 2013). Given this diversity of

function, it is not surprising that ncRNAs have been implicated in various process of multiple diseases including Alzheimer's, schizophrenia, autism spectrum disorder, Parkinson's, Angelman's syndrome and Huntington's (Skene and Grant, 2016). Besides data on micro RNAs (miRNAs), (Miranda et al., 2010; Lewohl et al., 2011; Nunez and Mayfield, 2012; Nunez et al., 2013; Gorini et al., 2013) the association of ncRNAs with excessive ethanol consumption is not clear, especially in terms of their potential roles in the risk for excessive ethanol consumption (Mayfield, 2017). ncRNAs that may have a role include the small nuclear RNAs (snRNAs; ~150nt) and the long non-coding RNAs (lncRNAs; > 200nt). To our knowledge, there are no reports linking the various classes of snRNAs to excessive ethanol consumption. The situation is different for the lncRNAs. The nuclear-enriched abundant transcript 2 (NEAT2), an lncRNA important in synaptogenesis (Bernard et al., 2010), is reported to be upregulated in the human alcoholic brain (Kryger et al., 2012). NEAT2, and other related lncRNAs have also been found to be upregulated in the brains of heroin and cocaine abusers (Albertson et al., 2004; Michelhaugh et al., 2011). Using a network-based approach, Farris and Mayfield (2014) identified 3 long intergenic ncRNAs (lincRNAs) associated with chronic alcoholism. Xu et al. (2015) used GWAS data to identify an anti-sense overlapping lncRNA that covered multiple loci associated with ADH genes as a potential risk variant for alcoholism.

In addition to the focus on RNA-Seq and ncRNAs, the current study asked whether selection had differential effects on male and female gene expression networks. There is ample evidence of interactions between alcohol use disorders and sex (biological factors) and gender (psychological, social, and cultural factors) (reviewed in (Erol and Karpyak, 2015). The analyses of the strain distribution patterns for DID in the LXS recombinant inbred mouse strains revealed a number of male-specific quantitative trait loci (QTLs), which indicated the locations of genes or groups of closely linked genes affecting DID (Vanderlinden et al., 2015). QTL analyses for ethanol preference have also found a number of sex-specific loci (Melo et al., 1996; Gill et al., 1998; Bice et al., 2006; Vendruscolo et al., 2006; Izidio et al., 2011). However, from the perspective of the transcriptome, our own work (e.g. Iancu et al. 2013a; Metten et al., 2014) and that of others (e.g. Mulligan et al., 2011) have either not considered sex as a key analysis variable and/or have only included males in the experimental design. Wilhelm et al. (2014) have provided perhaps the most compelling evidence for sex-specific effects of alcohol on the brain transcriptome. These authors found a strong effect of sex on gene expression profiling patterns during chronic ethanol intoxication and peak withdrawal. During peak withdrawal a pro-inflammatory transcriptional phenotype was observed in females while in males there was a suppression of immune signaling. These data are consistent with the growing evidence for gene network patterns and gene network changes that are sex-specific (e.g. Oldham et al., 2008; van Nas et al., 2009; Wong et al., 2014; Irizar et al., 2014).

Methods

Animals

For gene expression analyses, selected generation 15 (S₁₅) HDID-2 mice were used; contemporary HS/NPT mice from the 72nd filial generation (G₇₂) were used as the controls.

Details concerning the selection of the HDID mice are found in (Crabbe et al., 2009, 2014; Barkley-Levenson and Crabbe, 2012a, 2014). Details about some genetic differences between the HDID-2 and HS/NPT mice are found in Iancu et al. (2013b). Animals were raised on a reversed light/dark cycle, with lights on between 9:30 PM and 9:30 AM. All animal care, breeding, and testing procedures were approved by the Institutional Animal Care and Use Committees at the Veterans Affairs Medical Center, Portland, OR, and the Oregon Health & Science University, Portland, OR.

Dissection of Tissue and Extraction of RNA

Naive mice (N=24/genotype/sex) were euthanized between 10 AM and 2 PM, brains removed and immediately frozen on dry ice. Frozen brains were dissected by hand under RNase-free conditions. Using the optic chiasm as caudal marker, a 2 mm coronal slice of brain tissue was isolated. Following the partial cone shape of the striatum, the dissection moved dorsal 1 mm, followed by a cut to the lateral boundary of the striatum, with a final cut following the lateral-ventral boundary (Figure 1). The isolated tissue was immediately placed into 1 ml of Trizol (Invitrogen, Carlsbad, CA) and a standard extraction protocol was followed according to manufacturer's instructions. Further details are found in (Zheng et al., 2015).

RNA-Seq

Library formation (Ribozero, stranded) and sequencing on an Illumina HiSeq 2000 were performed at the Oregon Health & Science University Massively Parallel Sequencing Shared Resource following the company specifications. Prior to library formation, RNA was prepared by removing adapter dimers; this removes RNA material of approximately 70 bp or less. Libraries were then formed for each subject and were constructed in one batch to minimize batch effects. Libraries were multiplexed four per lane, balanced for group and sex, yielding approximately 30 million total reads per sample. Sequencing required 2 flowcells; no flowcell batch effect was detected. FastQC was used to inspect the quality of the raw sequence data. All samples were reviewed to ensure that there was the expected enrichment in striatal genes (cortical samples are used for the comparator). Reads (100bp, single end) were then aligned to mouse reference genome mm10 using STAR version 2.5.3b (Dobin et al., 2013) with default parameters except for the following:

outFilterMismatchNmax = 3, outFilterScoreMinOverLread = 0.33, outFilterMismatchNoverLmax = 0.03 and outFilterMultimapNmax = 1. Using the Bedtools suite version 2.17.0 (Quinlan, 2014) and the combination of the Mus_musculus.GRCm38.85 Ensembl GTF annotation file and NONCODE2016_mouse_mm10_lncRNA.gtf annotation file from the NONCODE database (<http://www.noncode.org>), the read counts were summarized at the gene level with the following parameters: -S (because the strand from our Illumina sequenced fastq files are provided in inverted form) and -split. To eliminate ambiguities due to overlapping genomic features, only reads that were located on non-overlapping exon portions were retained for further analysis. Data were then imported into the R application environment (version 3.4.3). Upper quartile normalization was performed using the edgeR Bioconductor package. Genes with at least 1 count per million reads (CPM) for at least half the samples were retained for further analysis.

Differential expression (DE), differential variability (DV) and differential wiring (DW)

Our main goal was to quantify transcriptional differences between the HS and HDID animals and to further identify sex specific changes versus changes that affect both males and females. The analysis strategy therefore needed to balance the need for sufficient samples for network construction, the likely presence of sex specific effects, and the benefits of having a consensus network acting as a common structure for mapping selection driven transcriptional changes. In light of these somewhat competing requirements, network construction and analysis was performed separately for males and females. However, within each sex the HS and HDID animals were pooled together to construct male and female “consensus” networks.

DE between the naïve and selected groups was determined using edgeR (version 3.18.1) (Robinson et al., 2010), with the option of “tagwise” dispersion; the threshold for significance was either unadjusted $p=0.01$ for network level integration or at adjusted $p=0.05$. Multiple testing adjustment was performed utilizing the SGoF procedure (Carvajal-Rodríguez et al., 2009). For gene DV (Ando et al., 2015; Mar et al., 2011), we utilized the “var.test” procedure in the R “stats” package. Additionally, we evaluated the “differential wiring” (DW) between gene pairs. This procedure evaluates the statistical significance of differences in pairwise correlations between genes. For all gene pairs, we first computed Pearson correlations separately in the naïve and selected groups (also separately in males and females). Gene pairs in which the difference in Pearson correlation between HS/NPT and HDID groups was > 0.5 were retained for further statistical evaluation by a permutation procedure. We then combined the naïve and selected animals and we repeatedly ($N=200$) randomly divided this joint group in two subgroups similar in size to the naïve/selected partition. Next, we computed and compared the differences in pairwise gene Pearson correlation between the random groups. Two genes were found to be differentially wired if the difference in Pearson correlation between the naïve and selected groups was unlikely to arise by chance ($p<0.01$) during the $N=200$ permutations. This procedure has been adapted with modifications from previous work in genomic (Gill et al., 2010) and neural imaging studies (S M Hadi Hosseini, 2012).

A statistically significant change in the Pearson correlation is also denoted as a changed network edge. We identified, for each gene, the number of changed edges and we inquired whether some genes have a disproportionate number of changing edges, utilizing the binomial test with the following parameters. The average incidence of changing edges (the rate of the binomial test) was computed by dividing the number of changed edges (at unadjusted $p<0.01$) by the total number of network edges. The number of trials (for each gene) was equal to the number of edges. The number of “successes” for a gene was equal to the number of significantly changed edges ($p<0.01$). Genes significantly enriched in changed edges were denoted as differentially wired. The union of DE, DV and DW genes are denoted as the “affected” set.

Coexpression network construction

An unsigned coexpression network was constructed by means of the WGCNA (version 1.51) (Iancu et al., 2012; Langfelder and Horvath, 2008). Briefly, Pearson correlation was

computed between all gene pairs and its absolute value was subsequently raised at a power $\beta=7$ to filter out low correlation values and emphasize strong ones. Two “consensus” networks were constructed utilizing all available male/female samples; modules were detected in each network by hierarchical clustering. Changes in the network structure between the HDID-2 and HS/NPT were evaluated as previously described (Iancu et al. 2013b). For each gene, total network connectivity was computed as the sum of all its network edges. Modular connectivity restricts the edges included to the gene’s own module. The functional significance of all modules was evaluated using Gene Ontology (GO) enrichment analysis using the GO-stats R package (Ashburner et al., 2000; Falcon and Gentleman, 2007). The background set of the GO analysis contained only genes included in the network construction. Because of the nested structure of the GO terms, a graph decorrelation procedure was used (Alexa et al., 2006). To implement a ranking procedure, we integrated differential network results at the module and gene summarization levels into a comprehensive gene screening procedure. Modules enriched in gene or edge changes were the primary focus of further annotations. At the individual gene level, we focused on module hubs, which we define as those in the 80th percentile in terms of intramodular connectivity. The GOrilla algorithm (Eden et al., 2009) was used to provide a visual representation of GO annotation enrichment and to examine annotation enrichment of selected groups of genes against a background set of the network genes included in the analysis. The EnrichR web tool (Chen et al., 2013; Kuleshov et al., 2016) was utilized to search for additional functional annotations and overlap with known biological factors, although in this case the background set is the whole transcriptome and not just the network genes. Modules were evaluated in enrichment in neuronal cell types by utilizing Fisher’s exact test and the data from Cahoy et al, 2008; p values were corrected utilizing the Bonferroni procedure and the number of modules.

Results

RNA-Seq

A ribosomal RNA depletion and stranded RNA-Seq protocol was used to examine the effects of selection on the ventral striatal transcriptome. RNA was checked for quality using the Bioanalyzer (Agilent). Quality assessment was based upon the RIN (RNA integrity number) which is a standardized measure of how intact RNA is. Libraries were prepared from the RNA using the TruSeq stranded protocol with ribosomal RNA depletion. Ribosomal RNA was depleted using rRNA removal beads. The remaining RNA, which included poly(A)+ RNA and noncoding RNA, was fragmented using divalent cations and heat. First strand synthesis was primed with random hexamers in the presence of Actinomycin D to prevent priming from any residual DNA template. The second strand was then synthesized using dUTP instead of dTTP to prevent initial PCR amplification from the second strand. The resulting double stranded cDNA was adenylated to enhance attachment to adapters, followed by ligation of standard indexed Illumina adapters. The resulting library was amplified using a limited number of cycles of polymerase chain reaction (PCR). The PCR product was separated from unincorporated nucleotides and adapter dimers using AMPure XP Beads (Agencourt). The library was profiled on the TapeStation (Agilent) for quality assessment. Library concentration was determined using real time PCR (Kapa

Biosystems and ABI/Fisher). Libraries were then sequenced with multiplexing on a HiSeq high throughput flow cell (Illumina).

Approximately 30 million reads were obtained for each sample. On average, 65% of the reads were uniquely aligned to the genome, 25% were aligned to multiple loci and 10% did not align. These results are comparable with our standard poly-A+ and stranded- protocol (e.g. Colville et al. 2017) where > 90% of the transcripts are uniquely aligned, 8% align to multiple loci and 2% are left without alignment to any location in the genome.

Non-polyadenylated transcripts tend to contain more repeat sequencings than polyadenylated transcripts. On average (males + females, HDID-2 + HS/NPT), 82% of the uniquely aligned reads mapped to protein coding genes and 14% mapped to non-coding genes (ncRNAs). Note that “ncRNAs” here refers to long intergenic ncRNAs (lincRNAs) and does not include intronic lincRNAs due to potential ambiguities in assigning the reads, since the vast majority of intronic lincRNAs overlap at least one exon. The remaining 4% of the uniquely aligned reads were distributed across several different categories; 1% of the reads were to small nuclear RNAs (scaRNA, snRNA and snoRNA).

The WGCNA (Zhang and Horvath 2005; Langfelder et al., 2011) was used to parse the RNA-Seq data into 34 coexpression modules (Iancu et al. 2012); the female and male data were analyzed separately, modules were detected and color coded independently so modules denoted with the same color in male and female networks are distinct (Supplemental Tables 1 and 2). Gene ontology (GO) annotations for the modules are found in Supplemental Tables 3 and 4. We detected module enrichment for genes associated with neurons, astrocytes and oligodendrocytes (Supplemental Table 7). It was observed that > 30% of the noncodes in the males and females were located within a single module, which happened to be denoted by the same brown color in both networks (colors are arbitrarily assigned). Of the ncRNAs in the male and female brown modules, 79% were in common.

The number of annotated protein coding genes in the brown modules was small (11 for females and 7 for males) and none were found in both. The brown modules in both the males and females were not of high quality ($Z < 4$; see Langfelder et al. 2011) but were of sufficient quality to distinguish the modules from the “grey” modules that contained the unclustered genes.

We further examined the network properties of noncodes and contrasted them with the properties of the protein coding genes. We found that the network connectivity (sum of edge strengths for each gene) was much lower for noncodes versus protein coding – about 5 times lower in both the male and female consensus networks ($p < 10^{-15}$). The Maximum Adjacency Rate (MAR) was also lower in the noncodes. In WGCNA the MAR signifies whether a node has many weak connections (low MAR) versus a few high strength edges (high MAR) and is formally defined for each gene as the sum of the squared adjacencies divided by the sum of the adjacencies. The lower MAR for noncodes held even after restricting the analysis to nodes in top 30% connectivity ($p < 10^{-5}$ for both networks). This indicates that the network and by implication the biological function of noncodes is dispersed through the transcriptome and is not restricted to only a few genes.

Differentially Expressed (DE) Genes

For females, 227 genes met the criteria ($FDR < 0.05$) as determined by EdgeR (Robinson et al., 2010) for DE between HDID-2 and HS/NPT animals. The “genes” affected included 23 noncodes, 188 protein coding and 4 snRNAs (Figure 2A and Supplemental Table 5). The gene showing the largest change was *ansnRNA*, *Gm22513*; for this gene, the ratio of HS/NPT to HDID-2 expression was > 100 fold. The noncode database (www.noncode.org) provides relative RNA-Seq expression data for 6 tissues: heart, hippocampus, liver, lung, spleen and thymus. Focusing on the three noncodes where the fold change was greatest for HS/NPT $>$ HDID-2 (NONMMU-G029007.2, -G029009.2, -G053859), relative expression was higher in the hippocampus compared to these 6 other tissues. All three of these noncodes were found in the saddlebrown gene expression module. Gene Ontology annotation of this module was modest but did include negative regulation of receptor recycling (Supplemental Table 3). The top three noncodes where HDID-2 expression was $>$ HS/NPT expression were (NONMMU-G031224.2, -G028977.2 and -G023935.2); these were distributed across three different coexpression modules (red, tan and brown).

Annotation was available for 194 of the DE genes in females (DE_f genes); however, there was no significant enrichment in any GO category (Function, Process or Component). Dividing the data into (–) and (+) DE_f genes (see below) did not result in the detection of additional enriched GO categories. The Enrichr tool (Chen et al., 2013; Kuleshov et al., 2016) revealed that the DE_f expressed genes were not enriched in transcription factor or miRNA binding sites. Modest but significant enrichment (adjusted $p < 0.05$) of the DE_f genes was found in 2 modules, red and saddlebrown (see above). The red module was enriched in genes associated with transcription factor activity, sequence-specific DNA binding ($FDR < 0.05$) and in genes associated with astrocytes (see Supplemental Tables 3 and 7). From the network co-expression analysis, a relative measure of intramodular connectivity is also found in Supplemental Table 1; genes with values ≥ 0.8 (top 20%) were considered module hubs. Relative intramodular connectivity was low for the DE_f genes but was significantly higher in the HDID-2 as compared to the HS/NPT sample (0.29 vs 0.13, $p < 0.02$). There were however, two hub nodes, *Fabp7* and *Gbp7*, which showed a marked change in connectivity, defined here a change in relative connectivity of > 0.5 .

For males, 1525 genes met the criteria ($FDR < 0.05$) for DE_m expression. 153 DE genes overlapped between the males and females, although not necessarily in the same direction. Only 10 of these genes were also among the genes detected as DE in HDID2 in our previous microarray study (Iancu et al. 2013b); these gene group consists of *Rpl29*, *Pop4*, *Emp1*, *Tesk1*, *Aebp1*, *Nudt2*, *Dpp7*, *Adam22*, *Entpd2*, and *Vcp*. The DE_m genes included 7 linc RNAs (Ensembl annotation), 496 noncodes, 954 protein coding genes and 9 snRNAs. (Figure 2B and Supplemental Table 6) The noncodes showing the most significant effects in the females were also the most significant in the males. The enrichment of noncodes in the DE_m sample was highly significant ($p < 10^{-72}$); the enrichment of these noncodes in the brown module (342/496) was also significant ($p < 10^{-15}$). Similar to the female data, the largest DE_m gene in terms of fold-change was the snRNA *Gm22513*.

The DE_m expressed genes were divided into negative [(–) DE_m] (HDID-2 $<$ HS/NPT) and positive [(+) DE_m] (HDID-2 $>$ HS/NPT) categories; there were 836 (–) DE_m genes and 689

(+) DE_m genes. 449 of the DE_m noncodes (91%) were (–) DE_m genes; the skewing of the noncode distribution was highly significant ($p < 10^{-90}$). The annotated (+) DE_m genes showed no significant GO enrichment (Eden et al. 2009) although there was a trend ($FDR < 0.2$) for an enrichment in cell adhesion molecules. Annotation of the (–) DE_m subset revealed significant enrichments in the Process, Function and Component categories (Supplemental Table 6). Key enrichments were observed for extracellular matrix (ECM) organization ($FDR < 4 \times 10^{-6}$), immune system process ($FDR < 2 \times 10^{-4}$), collagen trimer ($FDR < 9 \times 10^{-16}$) and plasma membrane part ($FDR < 3 \times 10^{-11}$). A graphical representation of the enrichments in the GO Component category is depicted in Figure 3. The (–) DE_m genes were modestly enriched in transcription factor (TF) binding sites for SND1 ($FDR < 0.05$); there was no significant enrichment for miRNA binding sites.

The annotated (–) DE_m expressed genes were significantly enriched (adjusted $p < 0.01$) in four of the gene co-expression modules: lightyellow, pink, tan and violet. To directly compare with annotations noted above, the GOrilla algorithm (Eden et al., 2009) was used for additional annotation of these four modules (Supplemental Table 4). We focus here on the tan and violet modules. The tan module was significantly enriched in annotations for cilium movement ($FDR 3 \times 10^{-12}$), extracellular region ($FDR 3 \times 10^{-12}$), dynein complex ($FDR < 7 \times 10^{-6}$) and neuronal genes ($p < 0.0001$). The violet module was significantly enriched in annotations for immune system process, ($FDR 3 \times 10^{-31}$), receptor binding ($FDR 6 \times 10^{-6}$), and extracellular region ($FDR 2 \times 10^{-14}$) but not a specific cell type. The genes associated with each of these categories are found in Supplemental Table 4. Eleven (–) DE_m genes were hub nodes that met the criteria for a marked change (see above) and eight were found in the violet module: *Adgre5*, *Anxa1*, *Fn1*, *Lgal*, *Lyz2*, *Myo1f*, *Pglyrp1* and *Vwf*. Overall, relative intramodular connectivity in the (–) DE_m genes was low but was significantly higher in the HS/NPT as compared to the HDID-2 (0.25 vs 0.20, $p < 2 \times 10^{-7}$).

Differentially Variable (DV) Genes

DV was calculated based on the F-test as implemented in the `var.test` function in R, following the general procedure outlined elsewhere (Ando et al., 2015; Mar et al., 2011) and modified as described in Methods; see also Colville et al. (2017). Note that DE genes were not included in the analysis. For females, 1498 genes showed a significant difference in variability ($FDR < 0.05$). The DV_f genes included 175 noncodes, 1282 protein coding and 26 snRNA genes (Figure 2C). Using the same convention as above for the DE genes, for females there were 80 (–) DV_f genes and 1418 (+) DV_f genes, i.e., there was a marked skewing of the data to greater variability in the HDID-2 animals. Of the 80 (–) DV_f genes, a significant enrichment was detected for cytoskeleton of presynaptic active zone ($FDR 3 \times 10^{-2}$) and axon part ($FDR 1 \times 10^{-2}$) (Supplemental Table 5); genes involved included *Bsn*, *Pclo*, *Syn1*, *Myoc*, *Nav1*, *Tubb4a*, *Cplx2* and *Ank3*. The (–) DV_f genes were significantly enriched in the lightyellow coexpression module ($p < 10^{-5}$). This module was characterized by annotations associated with regulation of transcription from RNA polymerase II promotor ($p < 6 \times 10^{-6}$), synapse ($p < 9 \times 10^{-5}$) and transcription regulatory region DNA binding ($p < 1 \times 10^{-6}$). Synaptic genes in this module included *Grik5*, *Grin1*, *Grin2b*, *Shank1*, *Sncb*, *Syn3*, *Syngap1* and *Vamp2*. On average among the (–) DV_f genes, relative intramodular connectivity was lower in the HDID-2 compared to the HS/NPT sample [0.20 vs 0.42] ($p <$

1×10^{-7}). There was no significant enrichment in (–) DV_f genes for specific TF or miRNA binding sites.

For the (+) DV_f genes there was a significant enrichment in annotations that included extracellular space (FDR 1×10^{-36}) plasma membrane part (FDR 1×10^{-7}), signaling receptor activity (FDR 1×10^{-5}) and extracellular matrix organization (FDR 1×10^{-8}) (Supplemental Table 5). The (+) DV_f genes were significantly (adjusted $p < 0.01$) enriched in 7 of the coexpression modules: darkturquoise, greenyellow, lightgreen, orange, steelblue, violet and white. Detailed GO annotations for the modules are found in Supplemental Table 3. The darkturquoise, orange and steelblue modules are enriched in annotations associated with the extracellular space, most significantly in the orange module ($p < 1 \times 10^{-21}$). Of the 7 modules, none were enriched in neuronal genes, 4 were enriched in genes associated with astrocytes (darkturquoise, greenyellow, violet and white) and 3 were enriched in genes associated with oligodendrocytes (darkturquoise, greenyellow and steelblue). On average among the (+) DV_f genes, relative intramodular connectivity increased in the HDID-2 compared to the HS/NPT sample (0.63 vs 0.37; $p < 10^{-143}$). There were 68 (+) DV_f genes that moved from non-hub status in the HS/NPT to hub status in the HDID-2 with a change in relative intramodular connectivity of >0.50 . These genes were primarily found in the greenyellow (11), lightgreen (18) and orange modules (18). There were only two noncodes in this group: NONMMUG062627.1 (brown module) and NONMMUG019998.2 (steelblue module). There was no significant enrichment in the (+) DV_f subset for specific TF or miRNA binding sites.

For males, 766 genes showed a significant difference in variability (FDR < 0.05). 82 genes overlapped between the male and female DV subsets. Included in the overlapping set were *Calb2*, *Gabrq*, *Nos1ap*, *Oxt*, *Pomc*, *Pvab*, *Slc6a11*, *Trh* and 14 noncodes. The male DV genes included 190 noncodes, 501 protein coding and 6 snRNA genes (Figure 2D). This subset was significantly enriched in noncodes ($p < 0.001$). The DV_m genes were distributed between 103 (–) DV_m and 663 (+) DV_m genes. For the (–) DV_m genes there were significant enrichments in GO categories associated with biological adhesion (FDR $< 1 \times 10^{-2}$) and extracellular part (FDR $< 2 \times 10^{-3}$). The annotated (–) DV genes were overexpressed in only one module, violet ($p < 5 \times 10^{-4}$). Annotations for this module are described above. As with the female data, relative intramodular connectivity decreased in the HDID-2 compared to the HS (0.19 vs 0.43; $p < 1 \times 10^{-10}$). Nine HS hub nodes showed a marked decrease in connectivity in the HDID-2 and 8 of these nodes were in the violet module. Four noncodes also showed a similar decrease in connectivity; all of these noncodes were in the brown module. There was no significant enrichment in the (–) DV_m subset for specific TF or miRNA binding sites.

For the (+) DV_m genes there were significant enrichments in GO categories that included modulation of synaptic transmission (FDR 7×10^{-4}), voltage gated cation channel activity (FDR 2×10^{-4}), plasma membrane part (FDR 3×10^{-7}) and synapse part (FDR 4×10^{-5}). Genes in the latter category included *Grin2a*, *Grin2b*, *Dlg4*, *Gabbr2*, *Grm2*, *Pdyn*, *Gabra1* and *Camk2a*. The two coexpression modules showing the greatest enrichment in (+) DV_m genes were the magenta ($p < 2 \times 10^{-8}$) and steelblue ($p < 3 \times 10^{-7}$) modules. Both modules were highly enriched in neuronal genes ($p < 10^{-4}$) (Supplemental Table 7). The magenta

module was enriched in annotations that included extracellular region ($\text{FDR} < 8 \times 10^{-11}$), cation channel activity ($\text{FDR} < 7 \times 10^{-5}$) and signal transduction ($\text{FDR} < 3 \times 10^{-5}$). The steelblue module was significantly enriched in annotations that included behavior ($\text{FDR} < 6 \times 10^{-5}$), neuron projection ($\text{FDR} < 2 \times 10^{-6}$) and RNA polymerase II TF activity and sequence-specific DNA binding ($\text{FDR} < 2 \times 10^{-4}$). Genes in the behavior annotation included *Gabra5*, *Prkca*, *Slc1a2*, *Htr1a*, *Npy1r*, *Scn8a* and *Adra1b*. Only four protein coding (+) DV_m genes showed a marked effect of selection on hub status: *Map3k14*, *Uso1*, *Mkap1a* and *Megf9*. Three noncodes also showed a marked change in hub status and one of these *A830039N20Rik* was highly expressed in the steelblue module. Overall, relative intramodular connectivity was increased in the HDID-2 compared to the HS/NPT (0.51 vs 0.32; $p < 4 \times 10^{-35}$). There was no significant enrichment in the (+) DV_m subset for specific TF or miRNA binding sites.

Differentially Wired (DW) Genes

DW (Colville et al. 2017) provides a measure of changing connectivity to the entire co-expression network, independent of the DE and DV genes. For females, 1307 genes met the criteria for DW; this subset included 388 noncodes, 806 protein coding and 60 snRNA genes (Figure 2E). The noncode ($p < 0.004$) and snRNA ($p < 0.0003$) genes were significantly enriched. The DW_f were enriched in annotations associated with the extracellular region ($\text{FDR} < 4 \times 10^{-4}$) and more modestly enriched in annotations associated with multicellular organismal process ($\text{FDR} < 3 \times 10^{-2}$) and tissue development ($\text{FDR} < 3 \times 10^{-2}$). The DW_f genes were enriched (adjusted p value < 0.001 or better) in four co-expression modules: black, darkolivegreen, lightyellow and midnightblue. The characteristics of the lightyellow module are described above. Details on the black and darkolivegreen modules are found in Supplemental Table 4. The midnightblue module was enriched in annotations associated with synaptic signaling ($\text{FDR} < 8 \times 10^{-6}$), transmembrane receptor activity ($\text{FDR} < 8 \times 10^{-3}$), plasma membrane ($\text{FDR} < 4 \times 10^{-6}$) and G-protein coupled receptor signaling pathways ($\text{FDR} < 2 \times 10^{-4}$). G-protein associated genes found in the midnightblue module included *Adora2*, *Drd2*, *Pde10a*, *Ppp1r1b*, *Rgs9*, & *Sstr5*. There were also 48 DW_fnoncodes in the midnightblue module, suggesting they have an association with synaptic function. Eighteen protein coding DW_f genes (and no noncode or snRNA genes) showed a marked change in node hub status: HDID-2 >> HS/NPT. These genes were widely distributed across the co-expression modules although 4 (*Usp36*, *Sertm1*, *Hsf4* and *Irgm1*) were found in the small violet module. There was no significant enrichment in the DW_f subset for specific TF or miRNA binding sites. For males, 706 genes met the criteria for DW; this subset included 171 noncodes, 491 protein coding and 11 snRNA (all snoRNA) genes (Figure 2F). The noncodes among the DW genes were significantly enriched ($p < 1 \times 10^{-12}$). These noncodes were not clustered in the brown module but were widely distributed across all of the co-expression modules. The DW_m genes were enriched in annotations associated with G-protein coupled receptor activity ($\text{FDR} < 1 \times 10^{-5}$) and the extracellular region ($\text{FDR} < 7 \times 10^{-4}$). Genes in the G-protein annotation included *Adra2b*, *Avpr1a*, *Crh2*, *Gpr1*, 3, 26, 151 & 162, *Hrh2*, *Mc4r*, *Npy1r*, *Npy2r*, and *Oprd1*. After correction for multiple comparisons there was no significant enrichment of the DW_m genes in any of the co-expression modules; however, there was a trend for enrichment ($p < 0.07$) in the violet module. There was no significant enrichment in the DW_m subset for specific TF or miRNA binding sites.

Male and females results overlap

The overlap between the genes affected in the males vs females, while relatively modest, was highly significant ($p < 10^{-16}$, Fisher test odds ratio > 4) (Figure 3A). The GO annotations displayed a stronger level of overlap ($p < 10^{-16}$, Fisher test odds ratio > 700) (Figure 3B, C). The identity of the common ontological categories included membrane, extracellular matrix and signaling (Figure 3D).

Discussion

The goal of the present study was a) to extend our understanding of how HDID selection affected the brain transcriptome, b) to contrast selection effects on males and females and c) to examine the selection relationship with non-coding RNAs. We utilized similar strategies to examine how selection for 2-bottle preference affected the transcriptome (Colville et al. 2017). Preference selection had marked effects on one network module significantly enriched in receptor signaling activity genes including *Chrna7*, *Grin2a*, *Htr2a*, and *Oprd1*. Selection also strongly affected the expression of cadherins and protocadherins, synaptic tethering molecules; further, it appeared that an lncRNA (*Gm26672*) was involved in the selection effects on protocadherin expression. Given that preference and DID are genetically partially distinct (Crabbe et al., 2011), it was expected and observed that there would be limited overlap between those results and what we have reported here. However, both selections had strong effects on gene clusters with membrane annotations.

There is now ample evidence that alcohol and other drugs of abuse can have marked effects on ECM constituents (reviewed in Lubbers et al., 2014 and Lasek, 2016). Ethanol has been shown, for example, to affect the brain expression of *tPA* (or *Plat*) (Pawlak et al., 2005; Bahi and Dreyer, 2012), *Mmp-9* (Wright et al., 2003), *Bcan* & *Ncan* (Coleman et al., 2014) and *Tsp2* & *Tsp4* (Risher et al., 2015). Indeed, there are some data showing that all elements of the brain ECM – the basement membrane (laminins and collagens), the interstitial ECM (collagens, fibronectins, tenascins) and the perineuronal nets (proteoglycans) – are affected by acute and/or chronic ethanol treatment (Lasek, 2016). The evidence that changes in the brain ECM are associated with the *risk* for developing an alcohol use disorder are less compelling. However, polymorphisms have been detected in *Mmp-9m*, *Tnc* & *Thn* in human alcoholics (Samochowiec et al., 2010; Zuo et al., 2012). GWAS have revealed a polymorphism in *Col6a3* associated with alcoholism (Adkins et al. 2017). Our data illustrate that selection for the HDID phenotype had broad effects on ECM associated genes in both males and females (Supplemental Table 8). The affected genes are associated with the basement membrane, the interstitial ECM, the perineuronal nets, and turnover of ECM constituents such as the matrix metalloproteinases and the tissue inhibitors of the matrix metalloproteinases. ECM has a role in a wide variety of functional processes, including synaptic plasticity (Lubbers et al., 2014). Here we note that the ECM can also have a significant role in the regulation of neuroimmune/neuroinflammatory responses that may be relevant to understanding excessive ethanol consumption (Cui et al., 2010; Khokha et al., 2013). For example, Seo et al. (2008) have found that (1) collagen induces an inflammatory response in microglia as evidenced by the production of nitric oxide, expression of inducible nitric oxide synthase, COX-2, CD40, and MMP-9; (2) DDR1 (a receptor tyrosine kinase) is

expressed in microglia and is phosphorylated by collagen treatment; and (3) collagen-induced microglial activation is abrogated by DDR1 blockade. Of related interest, Mulligan et al. (2006) observed that that brain *Ddr1* is differentially expressed between naïve preferring and non-preferring mouse selected lines. *Ddr1* is also differentially expressed between naïve P and NP rats ($P > NP$) (R. Bell, personal communication). The link between the ECM and immune responses is also evidenced in the male violet coexpression module that was strongly affected by selection. The violet module was highly enriched in genes associated with the extracellular region ($FDR < 2 \times 10^{-14}$) and immune system process ($FDR < 3 \times 10^{-31}$).

Although the HDID-2 males and females do not differ phenotypically other than females showing somewhat higher consumption than males (Crabbe et al., 2014), the current study focused on whether the transcriptional changes associated with selection differed between males and females. As noted above, in both sexes, selection affected large clusters of ECM genes; however, the effects on the ECM genes were in some cases very different, and further, the statistic for detecting the changes, e.g., DE vs DV, reflected the often very different ways in which the sexes' transcriptomes responded. The underlying mechanisms associated with these differences are not clear but the data provide an important lesson, namely that relying on a single index to assess female-male differences can be an inadequate approach. We recognize that the female/male differences detected may simply be a sampling error and/or reflect insufficient statistical power. However, we note that while there are differences in expression variances between males and females, overall the variance in males and females was similar. Thus, we conclude that our observation that there were significantly fewer DE genes in females as compared to males is not simply the result of a variance expansion in the female samples.

In addition to the effects on the ECM, HDID selection had broad effects in both females and males on genes associated with synaptic signaling. There were however some sex differences, albeit somewhat subtle. For example, in the male network we detected a number of genes with significant DW between HDID-2 and NPT. This group included a number of G-protein coupled receptors such as *Adra2b*; *Avpr1a*; *Crh2*; *Gpr1*, 3, 26, 151 & 162; *Hrh2*; *Mc4r*; *Npy1r*; *Npy2r*; and *Oprd1*. Several of these genes have been associated with excessive ethanol consumption (Thiele and Navarro, 2014; Barkley-Levenson and Crabbe, 2012b) and in particular we note the NPY receptors. When examining the female network, the genes above were not DW, but a different group of G-protein coupled receptors (and related genes) did appear affected (e.g., *Adora2*, *Drd2*, *Pde10a*, and *Ppp1r1b*). Each of these genes has been significantly linked to alcohol and substance abuse (Celorrio et al., 2016; Logrip and Zorrilla, 2014). We conclude that in females and males, selection affected the same G-protein coupled pathways but the identity of the individual DW genes differed.

One goal of the current study was to determine to what extent ncRNAs were associated with selection and here the focus was on the lincRNAs. We detected > 2000 lincRNAs expressed above threshold. Repeatedly, we observed in both males and females that these lincRNAs were significantly enriched in the gene clusters affected by selection. In some cases these affected lincRNAs were highly enriched in the brown modules that were similar in composition in both males and females. For both sexes, there were only a handful of protein

coding genes in the brown modules and these were insufficient to provide module annotation. However, there are examples of annotation by association. For example, the midnightblue module in females was strongly affected by selection and contained 48 lincRNAs; the module was enriched in annotations associated with synaptic signaling, transmembrane receptor activity and plasma membrane. Although the lincRNAs were strongly affected by selection, there were relatively few examples of where these affected lincRNAs were hub nodes; overall protein coding genes formed the majority of the transcripts with >1 CPM and had significantly higher connectivity. However, there are still examples of high connectivity ncRNAs, albeit with lower MAR, indicating connections with numerous other nodes as opposed to simply being attached to a (possibly genetically fixed) close gene. We conclude that a relatively small minority of ncRNAs play significant transcriptional network roles. An example of a lincRNA hub node that was affected by selection and was highly expressed in the male steelblue module was *A830039N20Rik*. The steelblue module was significantly enriched annotations that included behavior, neuron projection and RNA polymerase II transcription factor activity. Based on our data we can conclude that multiple lincRNAs are associated with the risk for developing the HDID phenotype. We also note that snRNAs were affected by selection. In fact the gene showing the largest change in expression (HS/NPT>>>HDID-2) was the snRNA *Gm22513*. The precise function of this snRNA is unknown.

Independent of selection, the data presented here illustrate an important point. The first is that the genes that show significant DE are only rarely coexpression hub nodes and are often found in the bottom quartile of intramodular connectivity, i.e., they are leaf nodes. From the network perspective, these leaf nodes are likely to be suitable biomarkers but manipulating these genes is not likely to affect the phenotype of interest. The observation that the DE genes are leaf nodes follows from the fact that strongly DE genes will present a bimodal distribution and therefore linear correlations utilized to construct the network will appear low, which results in decreased connectivity. We circumvented these potential difficulties by separately evaluating DV and DW for all network genes. Variability increases in the HDID versus the NPT in many cases, even though some genes will fixate in the HDID. However, a majority of genes will only be partially fixated. If the “fixation point” differs from the center of the original distribution, then genes not fixated but in the intermediate stage will have a bimodal distribution with increased variance.

Mulligan et al. (2006), in the first meta-analysis of transcriptional changes associated with a behavioral trait, detected > 3000 genes associated with the risk for excessive consumption in the mouse ethanol preference (2-bottle choice) paradigm. The data were parsed into pathways and several pathways were highly enriched in the DE genes, e.g., MAPK/ERK1,2 signaling. However, despite this reduction in complexity, the risk factors remained numerous. Over the past dozen years, the complexity has not resolved and in fact one could argue has increased (see, e.g., Colville et al. 2017). Although we know less about the transcriptional features associated with the HDID phenotype, the situation appears to be essentially the same as for preference, i.e., multiple systems are affected (e.g. Mulligan et al. 2011; Iancu et al. 2013b). The current study similarly detects transcriptional changes between HDID and HS/NPT that are widely dispersed across multiple systems. However, we achieve a level of specificity based on the relatively small set of gene ontologies that are

enriched in affected genes; importantly several of these are independently detected in both males and females. In conclusion, our results support three main observations. First, and as with preference selection, HDID selection is associated with marked changes in genes that have membrane annotations. Second, new data are provided that the risk for excessive consumption is associated with a reorganization of the ECM. These data complement the observation that the ECM is affected by excessive consumption (Lasek, 2016). Third, the data clearly illustrate the advantages of analyzing the male and female data separately. Although several ontologies identified were the same in males and females, the individual gene identities and the nature of the gene level changes (DE, DV or DW) were strikingly different.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was supported in part by AA13519, AA 13484, AA10760, AA020245 and a grant from the US Department of Veterans Affairs. All authors confirm that they have no conflict of interests to declare with this submission.

References

- Adkins AE, Hack LM, Bigdeli TB, Williamson VS, McMichael GO, Mamdani M, Edwards AC, Aliev F, Chan RF, Bhandari P, Raabe RC, Alaimo JT, Blackwell GG, Moscatti A, Poland RS, Rood B, Patterson DG, Collaborative Study of the Genetics of Alcoholism Consortium, Walsh D; Whitfield JB, Zhu G, Montgomery GW, Henders AK, Martin NG, Heath AC, Madden PAF, Frank J, Ridinger M, Wodarz N, Soyka M, Zill P, Ising M, Nöthen MM, Kiefer F, Rietschel M, German Study of the Genetics of Addiction Consortium, Gelernter J, Sherva R, Koesterer R, Almasy L, Zhao H, Kranzler HR, Farrer LA, Maher BS, Prescott CA, Dick DM, Bacanu SA, Mathies LD, Davies AG, Vladimirov VI, Grotewiel M, Bowers MS, Bettinger JC, Webb BT, Miles MF, Kendler KS, Riley BP. 2017 Genomewide Association Study of Alcohol Dependence Identifies Risk Loci Altering Ethanol-Response Behaviors in Model Organisms. *Alcohol Clin Exp Res*. 5;41(5):911–928. doi: 10.1111/acer.13362. Epub 2017 Mar 30 PMID: 28226201. [PubMed: 28226201]
- Albertson DN, Pruetz B, Schmidt CJ, Kuhn DM, Kapatos G, Bannon MJ, 2004 Gene expression profile of the nucleus accumbens of human cocaine abusers: evidence for dysregulation of myelin. *J. Neurochem* 88, 1211–1219. [PubMed: 15009677]
- Ando T, Kato R, Honda H, 2015 Differential variability and correlation of gene expression identifies key genes involved in neuronal differentiation. *BMC Syst. Biol* 9. doi:10.1186/s12918-015-0231-6
- 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, 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet* 25, 25–29. doi:10.1038/75556 [PubMed: 10802651]
- Bahi A, Dreyer J-L, 2012 Involvement of tissue plasminogen activator “tPA” in ethanol-induced locomotor sensitization and conditioned-place preference. *Behav. Brain Res* 226, 250–258. doi: 10.1016/j.bbr.2011.09.024 [PubMed: 21945298]
- Barkley-Levenson AM, Crabbe JC, 2014 High drinking in the dark mice: a genetic model of drinking to intoxication. *Alcohol Fayettev. N* 48, 217–223. doi:10.1016/j.alcohol.2013.10.007
- Barkley-Levenson AM, Crabbe JC, 2012a Ethanol drinking microstructure of a high drinking in the dark selected mouse line. *Alcohol. Clin. Exp. Res* 36, 1330–1339. doi:10.1111/j.1530-0277.2012.01749.x [PubMed: 22524154]

- Barkley-Levenson AM, Crabbe JC, 2012b Bridging Animal and Human Models. *Alcohol Res. Curr. Rev* 34, 325–335.
- Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdain L, Couplier F, Triller A, Spector DL, Bessis A, 2010 A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* 29, 3082–3093. doi:10.1038/emboj.2010.199 [PubMed: 20729808]
- Bice PJ, Foroud T, Carr LG, Zhang L, Liu L, Grahame NJ, Lumeng L, Li T-K, Belknap JK, 2006 Identification of QTLs influencing alcohol preference in the High Alcohol Preferring (HAP) and Low Alcohol Preferring (LAP) mouse lines. *Behav. Genet* 36, 248–260. doi:10.1007/s10519-005-9019-6 [PubMed: 16482403]
- Cahoy John D., Emery Ben, Kaushal Amit, Foo Lynette C., Zamanian Jennifer L., Christopherson Karen S., Xing Yi, et al. 2008 “A Transcriptome Database for Astrocytes, Neurons, and Oligodendrocytes: A New Resource for Understanding Brain Development and Function.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28 (1): 264–78. 10.1523/JNEUROSCI.4178-07.2008. [PubMed: 18171944]
- Carvajal-Rodríguez A, de Uña-Alvarez J, Rolán-Alvarez E, 2009 A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests. *BMC Bioinformatics* 10, 209. doi:10.1186/1471-2105-10-209 [PubMed: 19586526]
- Celorrio D, Muñoz X, Amiano P, Dorronsoro M, Bujanda L, Sánchez M-J, Molina-Montes E, Navarro C, Chirlaque MD, MaríaHuerta J, Ardanaz E, Barricarte A, Rodriguez L, Duell EJ, Hijona E, Herreros-Villanueva M, Sala N, Alfonso-Sánchez MA, de Pancorbo MM, 2016 Influence of Dopaminergic System Genetic Variation and Lifestyle Factors on Excessive Alcohol Consumption. *Alcohol Alcohol. Oxf. Oxf* 51, 258–267. doi:10.1093/alcac/agv114
- Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma’ayan A, 2013 Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128. doi:10.1186/1471-2105-14-128 [PubMed: 23586463]
- Coleman LG, Liu W, Oguz I, Styner M, Crews FT, 2014 Adolescent binge ethanol treatment alters adult brain regional volumes, cortical extracellular matrix protein and behavioral flexibility. *Pharmacol. Biochem. Behav* 116, 142–151. doi:10.1016/j.pbb.2013.11.021 [PubMed: 24275185]
- Colville AM, Iancu OD, Oberbeck DL, Darakjian P, Zheng CL, Walter N.a. R., Harrington CA, Searles RP, McWeeney S, Hitzemann RJ, 2017 Effects of selection for ethanol preference on gene expression in the nucleus accumbens of HS-CC mice. *Genes Brain Behav.* 16, 462–471. doi: 10.1111/gbb.12367 [PubMed: 28058793]
- Crabbe JC, Harris RA, Koob GF, 2011 Preclinical studies of alcohol binge drinking. *Ann. N. Y. Acad. Sci* 1216, 24–40. doi:10.1111/j.1749-6632.2010.05895.x [PubMed: 21272009]
- Crabbe JC, Metten P, Belknap JK, Spence SE, Cameron AJ, Schlumbohm JP, Huang LC, Barkley-Levenson AM, Ford MM, Phillips TJ, 2014 Progress in a Replicated Selection for Elevated Blood Ethanol Concentrations in HDID Mice. *Genes Brain Behav.* 13, 236–246. doi:10.1111/gbb.12105 [PubMed: 24219304]
- Crabbe JC, Metten P, Rhodes JS, Yu C-H, Brown LL, Phillips TJ, Finn DA, 2009 A line of mice selected for high blood ethanol concentrations shows drinking in the dark to intoxication. *Biol. Psychiatry* 65, 662–670. doi:10.1016/j.biopsych.2008.11.002 [PubMed: 19095222]
- Cui P, Lin Q, Ding F, Xin C, Gong W, Zhang L, Geng J, Zhang B, Yu X, Yang J, Hu S, Yu J, 2010 A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* 96, 259–265. doi:10.1016/j.ygeno.2010.07.010 [PubMed: 20688152]
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR, 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl* 29, 15–21. doi:10.1093/bioinformatics/bts635
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z, 2009 GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48. doi: 10.1186/1471-2105-10-48 [PubMed: 19192299]
- Erol A, Karpyak VM, 2015 Sex and gender-related differences in alcohol use and its consequences: Contemporary knowledge and future research considerations. *Drug Alcohol Depend.* 156, 1–13. doi:10.1016/j.drugalcdep.2015.08.023 [PubMed: 26371405]

- Falcon S, Gentleman R, 2007 Using GStats to test gene lists for GO term association. *Bioinforma. Oxf. Engl* 23, 257–258. doi:10.1093/bioinformatics/btl567
- Farris SP, & Mayfield RD 2014 RNA-Seq Reveals Novel Transcriptional Reorganization in Human Alcoholic Brain. *International Review of Neurobiology*, 116, 275–300. doi.org/10.1016/B978-0-12-801105-8.00011-4 [PubMed: 25172479]
- Franklin K & Paxinos G 2007 *The Mouse Brain in Stereotaxic Coordinates*, 3rd Edition. Elsevier, Oxford, UK.
- Gill K, Desaulniers N, Desjardins P, Lake K, 1998 Alcohol preference in AXB/BXA recombinant inbred mice: gender differences and gender-specific quantitative trait loci. *Mamm. Genome Off. J. Int. Mamm. Genome Soc* 9, 929–935.
- Gill R, Datta S, Datta S, 2010 A statistical framework for differential network analysis from microarray data. *BMC Bioinformatics* 11, 95. doi:10.1186/1471-2105-11-95 [PubMed: 20170493]
- Gorini G, Nunez YO, Mayfield RD, 2013 Integration of miRNA and protein profiling reveals coordinated neuroadaptations in the alcohol-dependent mouse brain. *PloS One* 8, e82565. doi: 10.1371/journal.pone.0082565 [PubMed: 24358208]
- Guennewig B, Cooper AA, 2014 The central role of noncoding RNA in the brain. *Int. Rev. Neurobiol* 116, 153–194. doi:10.1016/B978-0-12-801105-8.00007-2 [PubMed: 25172475]
- Hitzemann R, Bottomly D, Darakjian P, Walter N, Iancu O, Searles R, Wilmot B, McWeeney S, 2013 Genes, behavior and next-generation RNA sequencing. *Genes Brain Behav.* 12, 1–12. doi:10.1111/gbb.12007 [PubMed: 23194347]
- Hitzemann R, Darakjian P, Walter N, Dan Iancu O, Searles R, McWeeney S, 2014 Introduction to Sequencing the Brain Transcriptome, in: *International Review of Neurobiology*. Elsevier, pp. 1–19. doi:10.1016/B978-0-12-801105-8.00001-1
- Iancu OD, Kawane S, Bottomly D, Searles R, Hitzemann R, McWeeney S, 2012 Utilizing RNA-Seq data for de novo coexpression network inference. *Bioinforma. Oxf. Engl* 28, 1592–1597. doi: 10.1093/bioinformatics/bts245
- Iancu OD, Oberbeck D, Darakjian P, Kawane S, Erk J, McWeeney S, Hitzemann R, 2013a Differential network analysis reveals genetic effects on catalepsy modules. *PloS One* 8, e58951. doi:10.1371/journal.pone.0058951 [PubMed: 23555609]
- Iancu OD, Oberbeck D, Darakjian P, Metten P, McWeeney S, Crabbe JC, Hitzemann R, 2013b Selection for drinking in the dark alters brain gene coexpression networks. *Alcohol. Clin. Exp. Res* 37, 1295–1303. doi:10.1111/acer.12100 [PubMed: 23550792]
- Irizar H, Muñoz-Culla M, Sepúlveda L, Sáenz-Cuesta M, Prada Á, Castillo-Triviño T, Zamora-López G, López de Munain A, Olascoaga J, Otaegui D, 2014 Transcriptomic profile reveals gender-specific molecular mechanisms driving multiple sclerosis progression. *PloS One* 9, e90482. doi: 10.1371/journal.pone.0090482 [PubMed: 24587374]
- Izídio GS, Oliveira LC, Oliveira LFG, Pereira E, Wehrmeister TD, Ramos A, 2011 The influence of sex and estrous cycle on QTL for emotionality and ethanol consumption. *Mamm. Genome Off. J. Int. Mamm. Genome Soc* 22, 329–340. doi:10.1007/s00335-011-9327-5
- Khokha R, Murthy A, Weiss A, 2013 Metalloproteinases and their natural inhibitors in inflammation and immunity. *Nat. Rev. Immunol* 13, 649–665. doi:10.1038/nri3499 [PubMed: 23969736]
- Kryger R, Fan L, Wilce PA, Jaquet V, 2012 MALAT-1, a non protein-coding RNA is upregulated in the cerebellum, hippocampus and brain stem of human alcoholics. *Alcohol Fayettev. N* 46, 629–634. doi:10.1016/j.alcohol.2012.04.002
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A, 2016 Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–97. doi:10.1093/nar/gkw377 [PubMed: 27141961]
- Langfelder P, Horvath S, 2008 WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559 [PubMed: 19114008]
- Langfelder P, Luo R, Oldham MC, Horvath S, 2011 Is my network module preserved and reproducible? *PLoS Comput. Biol* 7, e1001057. doi:10.1371/journal.pcbi.1001057 [PubMed: 21283776]

- Lasek AW, 2016 Effects of Ethanol on Brain Extracellular Matrix: Implications for Alcohol Use Disorder. *Alcohol. Clin. Exp. Res* 40, 2030–2042. doi:10.1111/acer.13200 [PubMed: 27581478]
- Lewohl JM, Nunez YO, Dodd PR, Tiwari GR, Harris RA, Mayfield RD, 2011 Up-regulation of microRNAs in brain of human alcoholics. *Alcohol. Clin. Exp. Res* 35, 1928–1937. doi:10.1111/j.1530-0277.2011.01544.x [PubMed: 21651580]
- Logrip ML, Zorrilla EP, 2014 Differential changes in amygdala and frontal cortex Pde10a expression during acute and protracted withdrawal. *Front. Integr. Neurosci* 8, 30. doi:10.3389/fnint.2014.00030 [PubMed: 24782725]
- Lubbers BR, Smit AB, Spijker S, van den Oever MC, 2014 Neural ECM in addiction, schizophrenia, and mood disorder. *Prog. Brain Res* 214, 263–284. doi:10.1016/B978-0-444-63486-3.00012-8 [PubMed: 25410362]
- Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, McGrath JJ, Quackenbush J, Wells CA, 2011 Variance of gene expression identifies altered network constraints in neurological disease. *PLoS Genet.* 7, e1002207. doi:10.1371/journal.pgen.1002207 [PubMed: 21852951]
- Mayfield RD, 2017 Emerging roles for ncRNAs in alcohol use disorders. *Alcohol Fayettev. N* 60, 31–39. doi:10.1016/j.alcohol.2017.01.004
- Melo JA, Shendure J, Pociask K, Silver LM, 1996 Identification of sex-specific quantitative trait loci controlling alcohol preference in C57BL/6 mice. *Nat. Genet* 13, 147–153. doi:10.1038/ng0696-147 [PubMed: 8640219]
- Mercer TR, Edwards SL, Clark MB, Neph SJ, Wang H, Stergachis AB, John S, Sandstrom R, Li G, Sandhu KS, Ruan Y, Nielsen LK, Mattick JS, Stamatoyannopoulos JA, 2013 DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat. Genet* 45, 852–859. doi:10.1038/ng.2677 [PubMed: 23793028]
- Metten P, Iancu OD, Spence SE, Walter NAR, Oberbeck D, Harrington CA, Colville A, McWeeney S, Phillips TJ, Buck KJ, Crabbe JC, Belknap JK, Hitzemann RJ, 2014 Dual-trait selection for ethanol consumption and withdrawal: genetic and transcriptional network effects. *Alcohol. Clin. Exp. Res* 38, 2915–2924. doi:10.1111/acer.12574 [PubMed: 25581648]
- Michelhaugh SK, Lipovich L, Blythe J, Jia H, Kapatos G, Bannon MJ, 2011 Mining Affymetrix microarray data for long non-coding RNAs: altered expression in the nucleus accumbens of heroin abusers. *J. Neurochem* 116, 459–466. doi:10.1111/j.1471-4159.2010.07126.x [PubMed: 21128942]
- Miranda RC, Pietrzykowski AZ, Tang Y, Sathyan P, Mayfield D, Keshavarzian A, Sampson W, Hereld D, 2010 MicroRNAs: master regulators of ethanol abuse and toxicity? *Alcohol. Clin. Exp. Res* 34, 575–587. doi:10.1111/j.1530-0277.2009.01126.x [PubMed: 20102566]
- Mulligan MK, Ponomarev I, Hitzemann RJ, Belknap JK, Tabakoff B, Harris RA, Crabbe JC, Blednov YA, Grahame NJ, Phillips TJ, Finn DA, Hoffman PL, Iyer VR, Koob GF, Bergeson SE, 2006 Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proc. Natl. Acad. Sci. U. S. A* 103, 6368–6373. doi:10.1073/pnas.0510188103 [PubMed: 16618939]
- Mulligan MK, Rhodes JS, Crabbe JC, Mayfield RD, Harris RA, Ponomarev I, 2011 Molecular profiles of drinking alcohol to intoxication in C57BL/6J mice. *Alcohol. Clin. Exp. Res* 35, 659–670. doi:10.1111/j.1530-0277.2010.01384.x [PubMed: 21223303]
- Nunez YO, Mayfield RD, 2012 Understanding Alcoholism Through microRNA Signatures in Brains of Human Alcoholics. *Front. Genet* 3, 43. doi:10.3389/fgene.2012.00043 [PubMed: 22514554]
- Nunez YO, Truitt JM, Gorini G, Ponomareva ON, Blednov YA, Harris RA, Mayfield RD, 2013 Positively correlated miRNA-mRNA regulatory networks in mouse frontal cortex during early stages of alcohol dependence. *BMC Genomics* 14, 725. doi:10.1186/1471-2164-14-725 [PubMed: 24148570]
- Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH, 2008 Functional organization of the transcriptome in human brain. *Nat. Neurosci* 11, 1271–1282. doi:10.1038/nn.2207 [PubMed: 18849986]
- Pawlak R, Melchor JP, Matys T, Skrzypiec AE, Strickland S, 2005 Ethanol-withdrawal seizures are controlled by tissue plasminogen activator via modulation of NR2B-containing NMDA receptors. *Proc. Natl. Acad. Sci. U. S. A* 102, 443–448. doi:10.1073/pnas.0406454102 [PubMed: 15630096]

- Quinlan AR, 2014 BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinforma* 47, 11.121–34. doi:10.1002/0471250953.bi1112s47
- Rhodes JS, Best K, Belknap JK, Finn DA, Crabbe JC, 2005 Evaluation of a simple model of ethanol drinking to intoxication in C57BL/6J mice. *Physiol. Behav* 84, 53–63. doi:10.1016/j.physbeh.2004.10.007 [PubMed: 15642607]
- Rhodes JS, Ford MM, Yu C-H, Brown LL, Finn DA, Garland T, Crabbe JC, 2007 Mouse inbred strain differences in ethanol drinking to intoxication. *Genes Brain Behav.* 6, 1–18. doi:10.1111/j.1601-183X.2006.00210.x [PubMed: 17233637]
- Risher M-L, Sexton HG, Risher WC, Wilson WA, Fleming RL, Madison RD, Moore SD, Eroglu C, Swartzwelder HS, 2015 Adolescent Intermittent Alcohol Exposure: Dysregulation of Thrombospondins and Synapse Formation are Associated with Decreased Neuronal Density in the Adult Hippocampus. *Alcohol. Clin. Exp. Res* 39, 2403–2413. doi:10.1111/acer.12913 [PubMed: 26537975]
- Robinson MD, McCarthy DJ, Smyth GK, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl* 26, 139–140. doi:10.1093/bioinformatics/btp616
- Hadi Hosseini SM, H. F, 2012 GAT: A Graph-Theoretical Analysis Toolbox for Analyzing Between-Group Differences in Large-Scale Structural and Functional Brain Networks. *PloS One* 7, e40709. doi:10.1371/journal.pone.0040709 [PubMed: 22808240]
- Samochowiec A, Grzywacz A, Kaczmarek L, Bienkowski P, Samochowiec J, Mierzejewski P, Preuss UW, Grochans E, Ciechanowicz A, 2010 Functional polymorphism of matrix metalloproteinase-9 (MMP-9) gene in alcohol dependence: family and case control study. *Brain Res.* 1327, 103–106. doi:10.1016/j.brainres.2010.02.072 [PubMed: 20197064]
- Seo M-C, Kim S, Kim S-H, Zheng LT, Park EK, Lee W-H, Suk K, 2008 Discoidin domain receptor 1 mediates collagen-induced inflammatory activation of microglia in culture. *J. Neurosci. Res* 86, 1087–1095. doi:10.1002/jnr.21552 [PubMed: 17969104]
- Skene NG, Grant SGN, 2016 Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front. Neurosci* 10. doi:10.3389/fnins.2016.00016
- Thiele TE, Navarro M, 2014 “Drinking in the dark” (DID) procedures: a model of binge-like ethanol drinking in non-dependent mice. *Alcohol Fayettev. N* 48, 235–241. doi:10.1016/j.alcohol.2013.08.005
- van Nas A, Guhathakurta D, Wang SS, Yehya N, Horvath S, Zhang B, Ingram-Drake L, Chaudhuri G, Schadt EE, Drake TA, Arnold AP, Lusis AJ, 2009 Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. *Endocrinology* 150, 1235–1249. doi:10.1210/en.2008-0563 [PubMed: 18974276]
- Vanderlinden LA, Saba LM, Bennett B, Hoffman PL, Tabakoff B, 2015 Influence of sex on genetic regulation of “drinking in the dark” alcohol consumption. *Mamm. Genome Off. J. Int. Mamm. Genome Soc* 26, 43–56. doi:10.1007/s00335-014-9553-8
- Vendruscolo LF, Terenina-Rigaldie E, Raba F, Ramos A, Takahashi RN, Mormède P, 2006 Evidence for a female-specific effect of a chromosome 4 locus on anxiety-related behaviors and ethanol drinking in rats. *Genes Brain Behav.* 5, 441–450. doi:10.1111/j.1601-183X.2005.00177.x [PubMed: 16923148]
- Wilhelm CJ, Hashimoto JG, Roberts ML, Sonmez MK, Wren KM, 2014 Understanding the addiction cycle: a complex biology with distinct contributions of genotype vs. sex at each stage. *Neuroscience* 279, 168–186. doi:10.1016/j.neuroscience.2014.08.041 [PubMed: 25194791]
- Wong RY, McLeod MM, Godwin J, 2014 Limited sex-biased neural gene expression patterns across strains in Zebrafish (*Danio rerio*). *BMC Genomics* 15, 905. doi:10.1186/1471-2164-15-905 [PubMed: 25326170]
- Wright JW, Masino AJ, Reichert JR, Turner GD, Meighan SE, Meighan PC, Harding JW, 2003 Ethanol-induced impairment of spatial memory and brain matrix metalloproteinases. *Brain Res.* 963, 252–261. [PubMed: 12560131]
- Xu K, Kranzler HR, Sherva R, Sartor CE, Almasy L, Koesterer R, Zhao H, Farrer LA, Gelernter J, 2015 Genomewide Association Study for Maximum Number of Alcoholic Drinks in European

- Americans and African Americans. *Alcohol. Clin. Exp. Res* 39, 1137–1147. doi:10.1111/acer.12751 [PubMed: 26036284]
- Zhang B, Horvath S, 2005 A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol* 4, Article17. doi:10.2202/1544-6115.1128
- Zheng CL, Wilmot B, Walter NA, Oberbeck D, Kawane S, Searles RP, McWeeney SK, Hitzemann R 2015 Splicing landscape of the eight collaborative cross founder strains. *BMC Genomics*, 16(1), 52. doi:10.1186/s12864-015-1267-0 [PubMed: 25652416]
- Zuo L, Gelernter J, Zhang CK, Zhao H, Lu L, Kranzler HR, Malison RT, Li C-SR, Wang F, Zhang X-Y, Deng H-W, Krystal JH, Zhang F, Luo X, 2012 Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol* 37, 557–566. doi:10.1038/npp.2011.229

Figure 1.

Coronal atlas section showing approximate caudal boundaries of dissection for ventral striatum. Heavy black lines indicate where tissue was sectioned (not shown is rostral boundary 2mm anterior); yellow shaded area depicts region of interest (modified from Franklin & Paxinos, 2007).

Figure 2.

Number and relative ratios of genes and noncodes related to selection for males and females. Orange: protein coding genes. Blue: noncodes. Gray: other transcripts (snRNAs, etc.). A,B: Differentially expressed (DE) transcripts. C,D: Differentially variable (DV) transcripts. E,F: Differentially wired (DW) genes.

Figure 3.

Male and female results show modest overlap in affected genes and strong overlap in GO annotations. A: Affected (DE, DV and DW) genes in males and females. B: Overlap in GO categories. C: Forest plot of enrichment of common genes and GO categories. D: Word cloud representation of the common GO categories.