bioRxiv preprint doi: https://doi.org/10.1101/752477; this version posted August 30, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Alcohol causes lasting differential transcription in *Drosophila* mushroom body neurons

Emily Petruccelli[a,b1]. Nicolas Ledru[a], Karla R. Kaun[a,1]

[a] Department of Neuroscience, Brown University, Providence, RI, USA, 02912
[b] Current address: Department of Biological Sciences, Southern Illinois University Edwardsville, Edwardsville, Illinois, USA, 62026
[1] Correspondence: karla_kaun@brown.edu or epetruc@siue.edu

Abstract

Repeated alcohol experiences can produce long-lasting memories for sensory cues associated with intoxication. These memories can ultimately trigger relapse in individuals recovering from alcohol use disorder (AUD). The molecular mechanisms by which alcohol changes memories to become long-lasting and inflexible remain unclear. New methods to analyze gene expression within precise neuronal cell-types can provide further insight towards AUD prevention and treatment. Here, we employed genetic tools in *Drosophila melanogaster* to investigate the lasting consequences of ethanol on transcription in memory-encoding neurons. *Drosophila* rely on mushroom body (MB) neurons to make associative memories, including memories of ethanol-associated sensory cues. Differential expression analyses found that distinct transcripts, but not genes, in the MB were associated with experiencing ethanol alone compared to forming a memory of an odor cue associated with ethanol. These findings reveal the dynamic and highly context-specific regulation of splicing associated with encoding behavioral experiences. Our data thus demonstrate that alcohol can have lasting effects on transcription and RNA processing during memory formation, and identify new transcript targets for future AUD and addiction investigation.

Keywords: alcohol, memory, RNA-seq, splicing, *Drosophila*

Introduction

Alcohol use disorder (AUD) impacts millions of individuals and constitutes one of the most serious public health problems in the world today (KENDLER *et al.* 2016; GRANT *et al.* 2017; CHENG *et al.* 2018). This chronic relapsing brain disorder can persist for extremely long periods of time regardless of alcohol abstinence. Relapse can be triggered by exposure to cues that predict alcohol availability, as they evoke memories of the drug's effects to result in cravings (JASINSKA *et al.* 2014; COURTNEY *et al.* 2016; GROEFSEMA *et al.* 2016; VALYEAR *et al.* 2017; CLEMENS AND HOLMES 2018; LOGGE *et al.* 2019). The complexity of alcohol's molecular actions and our limited understanding of how these actions change distinct neurons in the brain's reward memory circuitry has prevented our understanding of the mechanisms underlying these maladaptive memories.

Unlike the majority of other abused drugs, ethanol does not act on a single molecular target, but instead affects a variety of molecules (NESTLER 2013; TRUDELL *et al.* 2014). Many of the molecules implicated in ethanol-induced behaviors have broad roles in regulating diverse processes such as cell signaling, transcription, and neuronal plasticity (RON AND BARAK 2016; ERICKSON *et al.* 2019). These roles are more often than not dependent on both cell-type and developmental stage. The complexity of these processes, and diversity of experimental approaches in research studies, often obscures ethanol's context-dependent molecular effects. However, recent advances in cell-type specific isolation and sequencing technology can reveal the precise consequences of alcohol exposure on gene expression.

The genetic tools available in the fruit fly *Drosophila melanogaster* provide the ability to define where, when, and how alcohol may be acting in the nervous system, including within memory circuitry. *Drosophila* demonstrate ethanol-induced hyperactivity, tolerance (WOLF *et al.* 2002), and consummatory preference (DEVINENI AND HEBERLEIN 2009). They also remember and prefer the experience of intoxication (KAUN *et al.* 2011; NUNEZ *et al.* 2018), and this process requires mushroom body (MB) neuron activity. The MB integrates both sensory odor information from olfactory projection neurons and valence (aversive/rewarding) information from dopamine neurons. Downstream MB output neurons then drive avoidance of, or approach toward, an odor cue (ASO *et al.* 2014).

Groundbreaking studies have demonstrated that long-term memory formation in many species requires both *de novo* transcription and protein synthesis (see reviews (BAILEY *et al.* 1996; ALBERINI AND KANDEL 2014; SWEATT 2016). Although long-term memory formation requires transcriptional changes (ALBERINI AND KANDEL 2014; UCHIDA AND SHUMYATSKY 2018), it remains unclear whether repeated ethanol experiences influence these processes. Similarly, whether the presentation of ethanol alone or ethanol paired with an odor induces the same transcriptional events is unknown. We hypothesized that ethanol exposures, alone or paired with an odor, would produce lasting transcriptomic changes within memory-associated MB neurons, thus altering memory formation.

Recent RNA-sequencing (RNA-seq) analysis in *Drosophila* has systematically identified the transcriptomic profiles of various cell-types within MB circuitry (SHIH *et al.* 2019), and examined gene expression changes in the context of associative memory formation (CROCKER *et al.* 2016; WIDMER *et al.* 2018). We used nuclear, MB-specific RNA-seq to better understand transcriptional changes after the formation of alcohol cue associative memories. We found that alcohol exposures caused lasting expression changes at the transcript, but not gene, level, and that altered transcript expression was distinct between alcohol alone and alcohol cue memory groups. This suggests that RNA processing

bioRxiv preprint doi: https://doi.org/10.1101/752477; this version posted August 30, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

provides a distinct molecular landscape to help encode specific behavioral experiences.

Materials and Methods

Fly Husbandry

Flies were raised on cornmeal agar food at 24°C, 70% humidity, and with a 14:10 Light:Dark cycle. Male 4-5 day-old flies were isolated under CO_2, given one day to recover, and then conditioned with specified paradigms detailed below. The fly lines used in this study are a pan-MB specific driver *MB010B* split Gal4 line (Bloomington Stock #68293)(Aso *et al.* 2014) and the 5x*UAS-unc84*-2x*GFP* 'INTACT' line (Henry *et al.* 2012). Experimental flies were heterozygous for transgenes.

Odor cue-induced ethanol memory

Odor cue-induced ethanol memory was performed as published previously (Kaun *et al.* 2011; Nunez *et al.* 2018; Petruccelli *et al.* 2018). A minor difference was the use of smaller training/testing vials with 20-30 flies/vial. Humidified ethanol vapor (90:60 EtOH:air), resulting in a moderate 13.8 ± 3 mM (0.01 g/dL) internal body ethanol concentration was used. Ethanol, odors (1:36 odorant:mineral oil with ethyl acetate or iso-amyl alcohol), or both were delivered to flies in perforated vials for 10 min, resulting in a moderate internal ethanol concentration (Kaun *et al.* 2011). Following a 50 min rest, second and third sessions were then performed. Vapor treatments were delivered to flies on 1% agar and supplemented with yeast pellets overnight before being sacrificed the next day by liquid nitrogen freezing. All behavioral experiments were performed in reciprocal to control for innate odor preferences.

For each biological replicate, 2,000 male fly heads were collected via liquid nitrogen freezing and sieve separation 24 hours after conditioning. Flies were conditioned with one of four exposure paradigms (Figure 1A). Biological replicates were performed for Air (n = 3), EtOH (n = 3), Odors (n = 4), and Trained (n = 4) conditions. Odors and Trained conditions were performed with reciprocal odor groups (n = 2) (Ethyl Acetate→Isoamyl Alcohol and Isoamyl Alcohol→Ethyl Acetate).

Isolation of mushroom body nuclei (INTACT procedure)

Isolation of Nuclei Tagged in specific Cell Types (INTACT) was adapted from the method described in (Pankova and Borst 2016) and pioneered in flies by (Henry *et al.* 2012). The INTACT method of extracting nuclear RNA for sequencing provides several advantages compared to a general RNA-seq approach. First, the ability to use flash-frozen tissue, in contrast to FACS or dissected tissue samples, allows for an accurate examination of the current transcriptional profile of genetically targeted cells. Second, nuclear RNA from neurons contributes to the integrity of the active transcriptome snapshot by minimizing contamination from mRNA stored in the cytoplasm, along dendrites, and within axons, while allowing for the detection of experience-dependent differential expression (Lacar *et al.* 2016). Finally, given evidence that mRNA handling in subcellular compartments has been implicated in the formation and storage of memory (Bramham and Wells 2007; Richter 2010; Shigeoka *et al.* 2016; Nakahata and Yasuda 2018; Biever *et al.* 2019), this approach provides a robust profile of the stable post-exposure transcriptome unencumbered by the diversity of whole cell RNA.

Frozen heads were homogenized in a Kontes glass homogenizer (Sigma, D9938-1SET) with 10 mL of dounce buffer (10 mM ß-glycerophosphate, 2 mM $MgCl_2$, 0.5% Igepal buffer) and homogenized for ~2 min with the loose A pestle. Homogenate was passed through a 190 µm nylon net filter (Small Parts, CMN-0185-C). The filter was subsequently rinsed with 2 mL of the same buffer before discarding. Homogenate was subsequently further homogenized gently ~6-7 times with the tight B pestle, passed through a 20 µm filter (Small Parts, F020N-12-C), brought up to 50 mL by adding sucrose buffer (10 mM ß-glycerophosphate, 2 mM $MgCl_2$, 25 mM KCl, 250 mM sucrose). Next, 300 µL of Dynabead A magnetic beads (Thermo, 10001D) were pre-incubated with rabbit α-GFP antibody (Thermo, G10362), and incubated with the homogenate for 30 min at 4°C with gentle agitation. Beads were then captured with a magnet for 15 min and washed 5 times with 600 µL of sucrose buffer for 5 min at 4°C with gentle agitation. Finally, RNA was extracted from bead-bound nuclei using TRIzol (Ambion, Life Technologies), resuspended in RNA-ase free water and DNA-ase treated according to manufacturer's instructions (Ambion DNA-Free Kit).

RNA-seq

RNA libraries were polyA enriched to identify mRNA transcripts ready for nuclear export and sequenced using a Hi-Seq 4000 (Illumina) machine at a depth of ~30 million single-end 1x50bp reads by GENEWIZ (South Plainfield, NJ). Read quality was assessed using FastQC-0.11.5 (Babraham Bioinformatics). Adapters were removed and trimmed using Trimmomatic-0.36 (Bolger *et al.* 2014). The 'new tuxedo suite' was used to further process reads (Pertea *et al.* 2016). Reads were aligned with HISAT2-2.0.5 (Pertea *et al.* 2016) to the Ensembl BDGP6_transcriptome reference (Dm6), modified to include the *unc-84-2xGFP* sequence present in our flies. Next samtools-1.3.1 (Li *et al.* 2009) and stringtie-1.3.3 (Pertea *et al.* 2016) were used to sort, merge, and quantify transcripts. Lastly, assembly and analysis of Ballgown objects (Pertea *et al.* 2016) and data visualization with ggplot2 was performed in RStudio-1.0.136 (Wickham 2009). No samples were discarded and default settings were used for the pipeline unless otherwise noted - command line code as follows:

#Trimmomatic:

```
TruSeq3-SE.fa:2:30:10:8:true LEADING:3 TRAILING:3
SLIDINGWINDOW:4:15 MINLEN:36
```

#HISAT2:

```
hisat2 -q -x <reference index> -U <input fastq file>
-S <output sam file> -p 50
```

#Samtools:

```
samtools view -Su alns.sam | samtools sort -o
alns.sorted -@ 4
```

#StringTie:

```
stringtie <input.sorted.bam> -b <path to Ballgown
tables> -o <output.gtf> -G <reference.gtf>
```

Statistical Analysis

Libraries were statistically analyzed in R to determine differential expression between treatment conditions. Raw p-values are shown, as well as corrected p-values (false discovery rate < 0.05). The freely available Ballgown package (Pertea *et al.* 2016) afforded general linear modeling statistics for differential expression analysis and data visualization tools. Libraries were normalized with Ballgown package default – sum of each library's log expression measurements below the 75th percentile (Frazee

et al. 2015). For intersectional analyses, hypergeometric statistics were performed on pairwise comparisons and considered different from chance at p-value < 0.05.

Results

RNA-sequencing library characterization

We first generated flies that express the nuclear-tagged GFP 'INTACT' transgene (*UAS-unc84-GFP*) (HENRY *et al.* 2012) in all MB neurons (*MB010B-Gal4*). Expression of INTACT was MB specific and did not to disrupt alcohol cue memory formation (Supplemental Figure 1). MB-INTACT flies were subjected to one of four treatment paradigms – Air, EtOH, Odors, or Trained and flash frozen the following day (Figure 1A). MB nuclei were isolated and RNA sequenced to a depth of around 30 million reads per sample (Supplemental Figure 2). Nearly all libraries showed >80% uniquely mapped reads and ~15 million FPKM (Fragments Per Kilobase of transcript per Million mapped reads) per sample (Supplemental Figure 2).

A distribution of mean expression across the 17560 annotated genes revealed a similar frequency distribution of gene expression across treatment conditions (Figure 1C). Across the entire dataset, 7984 genes (45%) showed little to no expression (<2 log2(FPKM+1), whereas 9576 genes (55%) had robustly detected levels of gene expression (≥2 log2(FPKM+1). This suggested that adult MB nuclei were actively transcribing roughly 55% of annotated genes in the genome at the time of isolation.

The BDGP6 Dm6 transcriptome is comprised of 17560 genes and 34741 transcript isoforms with 10010 genes uniquely represented by a single transcript isoform. According to the current Flybase (flybase.org) annotation information, the Dm6 contains roughly 85% protein-coding isoforms, 2% noncoding RNAs, 1.5% microRNAs, 0.9% tRNAs, 0.8% snoRNAs, and 9.7% other/non-annotated biotype sequences. When pooling all expression counts across treatments, our data was comprised of 92% protein-coding genes/isoforms, 0.9% noncoding RNAs, 0.06% microRNAs, 0.06% tRNAs, 0.05% snoRNAs, and 6.3% other/non-annotated sequences (Figure 1D). Together the characterization of our RNA-sequencing libraries suggests that adult MB neurons are actively transcribing just over half of the genes in the genome and that there is an abundance of protein-coding genes detected to allow for interesting comparisons between treatment conditions.

Expression of *a priori* and novel gene lists across treatments

Since our RNA-seq approach was restricted to MB nuclei, we expected to see prominent expression of genes previously associated with MB function. Using the 'Vocabularies' search function on FlyBase (flybase.org), we acquired a list of genes associated with the 'adult mushroom body' (FBbt:00003684) including *Octopamine receptor in mushroom bodies* (*Oamb*), *Protein kinase cAMP-dependent regulatory subunit type 2* (*Pka-R2*), *short Neuropeptide F precursor* (*sNPF*), *Vesicular Acetylcholine Transporter* (*VAChT*), *portabella* (*prt*), *Dopamine Ecdysone Receptor* (*DopEcR*), etc. As expected, the majority of these MB-associated genes showed robust mean expression (>2 log2(FPKM+1)) across our treatment libraries (Figure 2A) thus supporting the MB specificity of the libraries.

Since the treatment conditions differed in the presence or absence of ethanol, we investigated how many ethanol-associated genes were expressed in MB nuclei. The acquired FlyBase gene list for 'behavioral response to ethanol' genes (FBbt:0048149) included

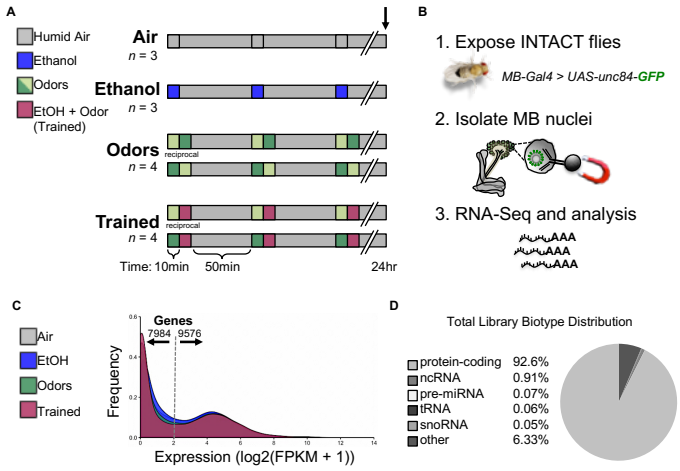


Figure 1. Mushroom Body INTACT paradigm and initial analysis. **A)** Paradigm depicting three spaced 10 minute exposures of 'Air', 'Ethanol', 'Odors', or 'Trained' (odor + ethanol) and sacrificed 24 hours later. Each biological replicate (n) was ~2000 males (Air: *n* = 3, Ethanol: *n* = 3, Odors: *n* = 4 (2 reciprocals), Trained: n = 4 (2 reciprocals). **B)** RNA-seq was performed on mushroom body (MB) nuclei of flies expressing the INTACT transgene (*MB10B-Gal4* > *UAS-unc-84-GFP*). Each fly brain contains ~4000 MB nuclei. **C)** Histogram of the mean gene expression in fragments per kilobase (FPKM) across treatment libraries. **D)** Percent distribution of biotypes according to current annotated Flybase 'feature' types across treatment libraries.

Sirtuin 1 (*Sirt1*), *Epidermal growth factor receptor* (*Egfr*), *homer* (*homer*), *Alcohol dehydrogenase* (*Adh*), *happyhour* (*hppy*), etc (Figure 2B). Forty-six of the fifty-five ethanol-associated genes showed robust expression (>2 log2(FPKM+1)) in the MB, but few genes showed noticeable differences between treatment conditions. This was not surprising, however, since most ethanol-associated genes were discovered using high-dose, acute ethanol treatment throughout the whole fly or heads, while this data shows gene expression 24 hours after repeated mild ethanol exposures specific to MB nuclei.

We next investigated the highest 20 transcribed genes across treatments (Figure 2C). As expected, we found the '*INTACTlabel*' (manually added to the Dm6 transcriptome) to be highly expressed across the libraries, demonstrating the robust nature of the *GAL4/UAS* binary expression system and further validating our MB isolation technique.

Aside from the INTACT label, most of the highly transcribed genes were RNA-associated genes (Figure 2Ci), which we posit to be a mix of transcripts from rough endoplasmic reticulum that was pulled down during nuclear isolation, as well as newly transcribed or nuclear-localized RNAs. Previous researchers using INTACT may not have observed these findings because ribosomal proteins were masked during transcriptome alignment (HENRY *et al.* 2012). Theoretically, nuclear RNA-seq should not require extensive ribosomal depletion and may allow for interesting rRNA expression analyses.

We then curated the highest 20 mean expressed genes after excluding RNA-associated genes (Figure 2Cii). These included genes encoding rhodopsin *NinaE* and rhodopsin inactivation protein (*Arr2*), translation elongation factors *EF1alpha48D* and *EF2*, a PKC inhibitor (*14-3-3ζ*), a vacuolar ATPase (*VHA16-1*), and cytoskeleton associated proteins *alphaTub84B*, *betaTub66D*,

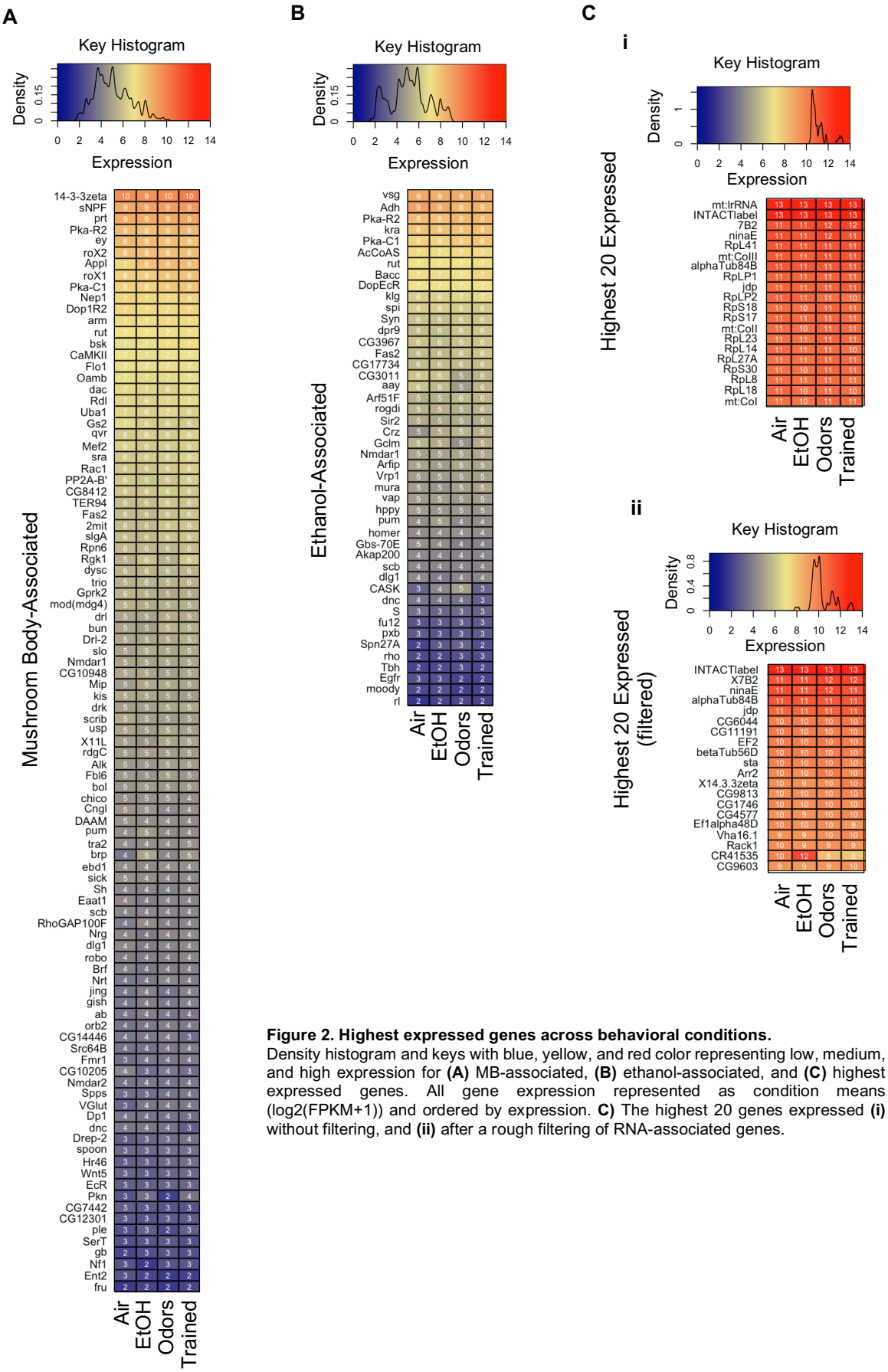
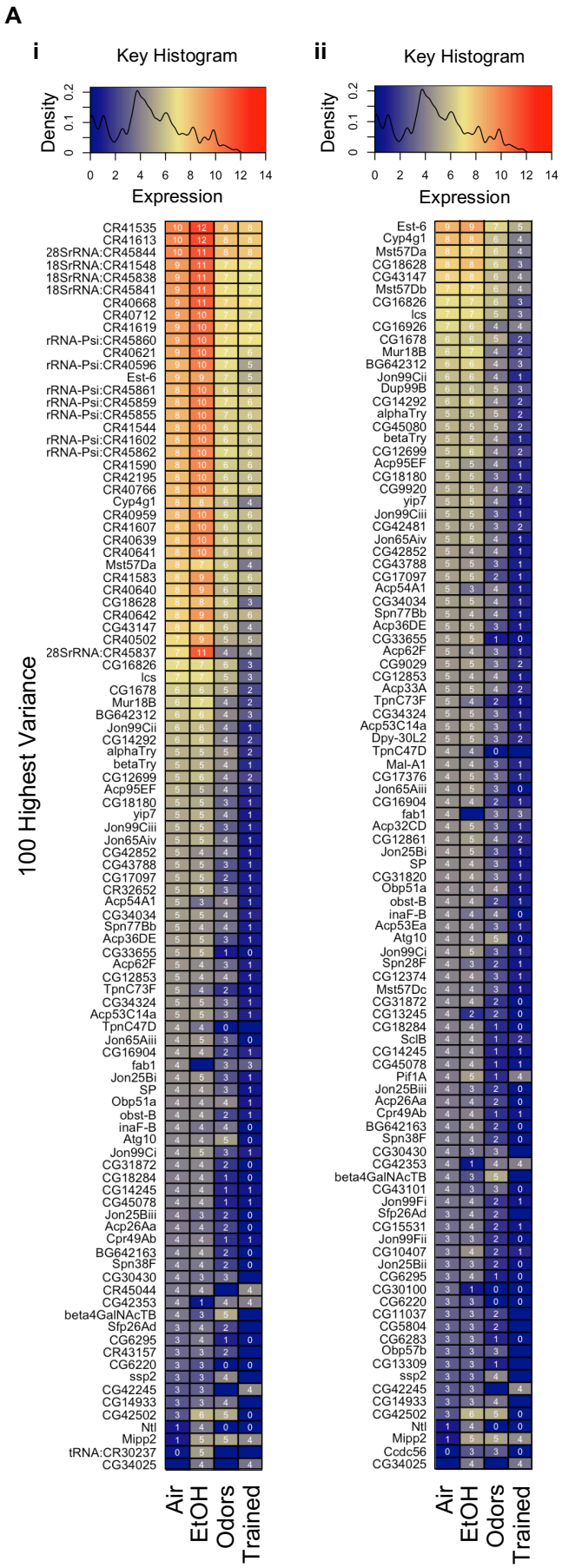


Figure 2. Highest expressed genes across behavioral conditions.
Density histogram and keys with blue, yellow, and red color representing low, medium, and high expression for **(A)** MB-associated, **(B)** ethanol-associated, and **(C)** highest expressed genes. All gene expression represented as condition means (log2(FPKM+1)) and ordered by expression. **C)** The highest 20 genes expressed **(i)** without filtering, and **(ii)** after a rough filtering of RNA-associated genes.



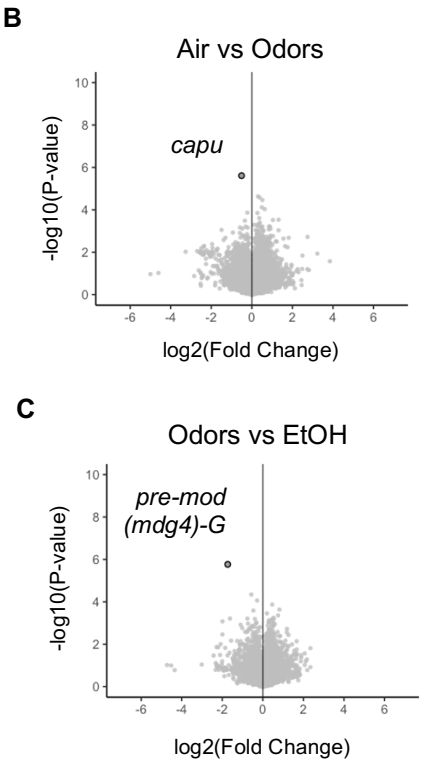


Figure 3. Top most variable and statistically significant genes between treatments.
A) Density histogram and keys with blue, yellow, and red color representing low, medium, and high expression of the 100 most variable genes – the squared deviation from the mean – **(i)** without filtering, **(ii)** after a rough filtering of RNA-associated genes. **B-C)** Volcano plot showing gene fold change (log2(fc) compared to the inverse of significance (-log10(p-value) (dark outline, FDR < 0.05). Both **(B)** Air vs Odors and **(C)** Odors vs EtOH had one statistically significant gene that changed expression.

Rack1 and others. Together, these descriptive analyses indicate that our libraries represent comparable MB datasets in which lasting experience-dependent molecular effects could be explored.

Experience-dependent differential gene expression

To uncover experience-dependent effects on transcription, we first examined genes that showed the greatest variance in expression (log transformed counts per million means) across treatments. This descriptive statistic reveals those genes with the greatest variability due to experience (variance is the square of the standard deviation from the mean). Among the 100 most variable genes, many ribosomal proteins and previously uncharacterized (unnamed CG#) genes were identified. Also included were *Esterase-6* (*Est-6*), *Cytochrome P450-4g1* (*Cyp4g1*), *la costa* (*lcs*), *Mucin related 18B* (*Mur18B*), *alpha and beta Trypsin* (*αTry*, *βTry*), *yippee interacting protein 7* (*yip7*), *fab1 kinase* (*fab1*), *Troponin C (TpnC73F and TpnC47D*), *Odorant binding protein 51a* (*Obp51a*), *obstructor-B* (*obst-B*), *inaF-B* (*inaF-B*), *Autophagy-related 10* (*Atg-10*) *short spindle 2* (*ssp2*), *neurotransmitter transporter-like* (*Ntl*), *Multiple inositol polyphosphate phosphatase 2* (*Mipp2*), and a number of *Jonah* peptidase transcripts (Figure 3Ai). Excluding RNA-associated genes further revealed *Maltase Ai*

(*Mal-A1*), *Sarcolamban B* (*SclB*), *Cuticular protein 49Ab* (*Cpr49Ab*) and *Coiled-coil domain containing 56* (*Ccdc56*) (Figure 3Aii).

Sex Peptide (*SP*), *Male-specific RNA 57Da and b* (*Msdt57Da, Msdt57Dd*), *BG642312*, *Serpins 77Bb* and *38F*, *Ductus ejaculatorius peptide 99B* (*Dup99B*), *Dpy-30-like 2* (*Dpy-30L2*), *PFTAIRE-interacting factor 1A* (*Pif1A*), *Seminal fluid protein 26Ad* (*Sfp26Ad*), and a number of *Accessory gland protein* (*Acp*) transcripts were also found, but may be a consequence of using male flies as these proteins are known to be produced by the male accessory glands. However, since heads were isolated prior to nuclear extraction, and the highly specific MB010B split-Gal4 line is not known to be expressed in the male accessory gland (Aso *et al.* 2014), our findings may be explained by a few contaminating whole fly bodies in some samples, or be unappreciated bona fide transcripts in MB neurons.

To determine differential expression between treatment conditions, we used 'Ballgown' (Pertea *et al.* 2016), a Bioconductor package. No genes or transcripts were removed from analyses and the adjusted p-value (a.k.a. q-value) was set to the standard false discovery rate (FDR) < 0.05. All pairwise condition contrasts were examined at both the gene and transcript level (Supplemental Figure 2). Only two genes were differentially expressed 24 hours after the various treatments, the actin filament nucleating protein *cappuccino (capu)* (Air vs Odors) (Figure 3B) and the nuclear DNA binding protein pre-*mod(mdg4)-G* (Odors vs EtOH) (Figure 3C).

Intersectional analysis of differential isoform expression

In contrast to gene-level analysis, there were more statistically significant differences found at the transcript level despite being subjected to more repeated measure correction. This suggests that there must be altered transcriptional start/stop site or RNA processing events that altered isoform diversity. In light of this finding, we focused on the differential expression of transcript isoforms between treatments.

Using Air treatment as a control, we compared the intersection of the top 200 most significantly differentially expressed (lowest p-values) isoforms from Air vs Odors (Figure 4Ai), Air vs EtOH (Figure 4Aii) and Air vs Trained (Figure 4Aiii). There was surprisingly little overlap in specific isoforms, including between 'Air vs EtOH' and 'Air vs Trained' (Figure 4B). This indicated that experiencing odors, ethanol intoxication, or making ethanol-odor memories produced distinct, molecularly separable changes in the lasting transcriptional state of MB nuclei. Using hypergeometric statistics, the number of overlapping isoforms was found to be statistically greater than would be expected by chance alone (Supplemental Figure 3). The extent of intersection was visualized with an upgraded Venn Diagram plot generated by an R package called 'UpSetR' (Figure 4B). Overlapping top p-value isoforms between treatment conditions are listed in Figure 4C.

Using the same logic just described, we compared the 200 isoforms with the largest fold changes in expression (Figure 4D). This approach highlights whether specific isoforms have greater dynamic range in expression in response to experiencing odors, ethanol, or ethanol-odor conditioning. Interestingly, there were more overlapping isoforms between pairwise comparisons than observed in the top p-value transcripts intersection analysis (Figure 4E). This finding suggests that particular isoforms in the MB neurons are more plastic than others to being changed in response to Odors, EtOH, and Trained treatments. Hypergeometric statistics again showed that the number of overlapping isoforms was statistically greater than would be expected by chance alone (Supplemental Figure 3). A subset of the overlapping top fold-changing isoforms between treatment conditions are listed in Figure 4F. The intersectional analyses between all pairwise library comparisons can be found in Supplemental Figure 4.

Enrichment of genes associated with differential isoforms

To determine which types of isoforms are enriched in the top 200 p-value comparisons, we performed DAVID Gene Ontology (GO) analysis on the associated gene IDs for each treatment - 167 Air, 175 Odor, 175, Ethanol, and 188 Trained genes. Using default DAVID settings, the top three GO annotations were identified (Supplemental Figure 5). Compared to Air controls, repeated Ethanol exposures produced statistically significant GO terms – 'Alternative splicing' (p-value = 0.0002), 'Phosphoprotein' (p-value = 0.041), and 'Coiled Coil' (p-value = 0.047). Interestingly, the response to Trained treatment also had 'Alternative splicing' (p-value = 0.062) as a high scoring enrichment category. It is important to note that the 'Alternative splicing' GO term does not refer to spliceosome genes, but rather to genes that are 'known to be alternatively spliced'. The enrichment analysis results further support the idea that there are experience-dependent changes in transcriptional start/stop site regulation or RNA processing, but not overall gene level changes. Perhaps future ontology databases will include isoform-specific categorization.

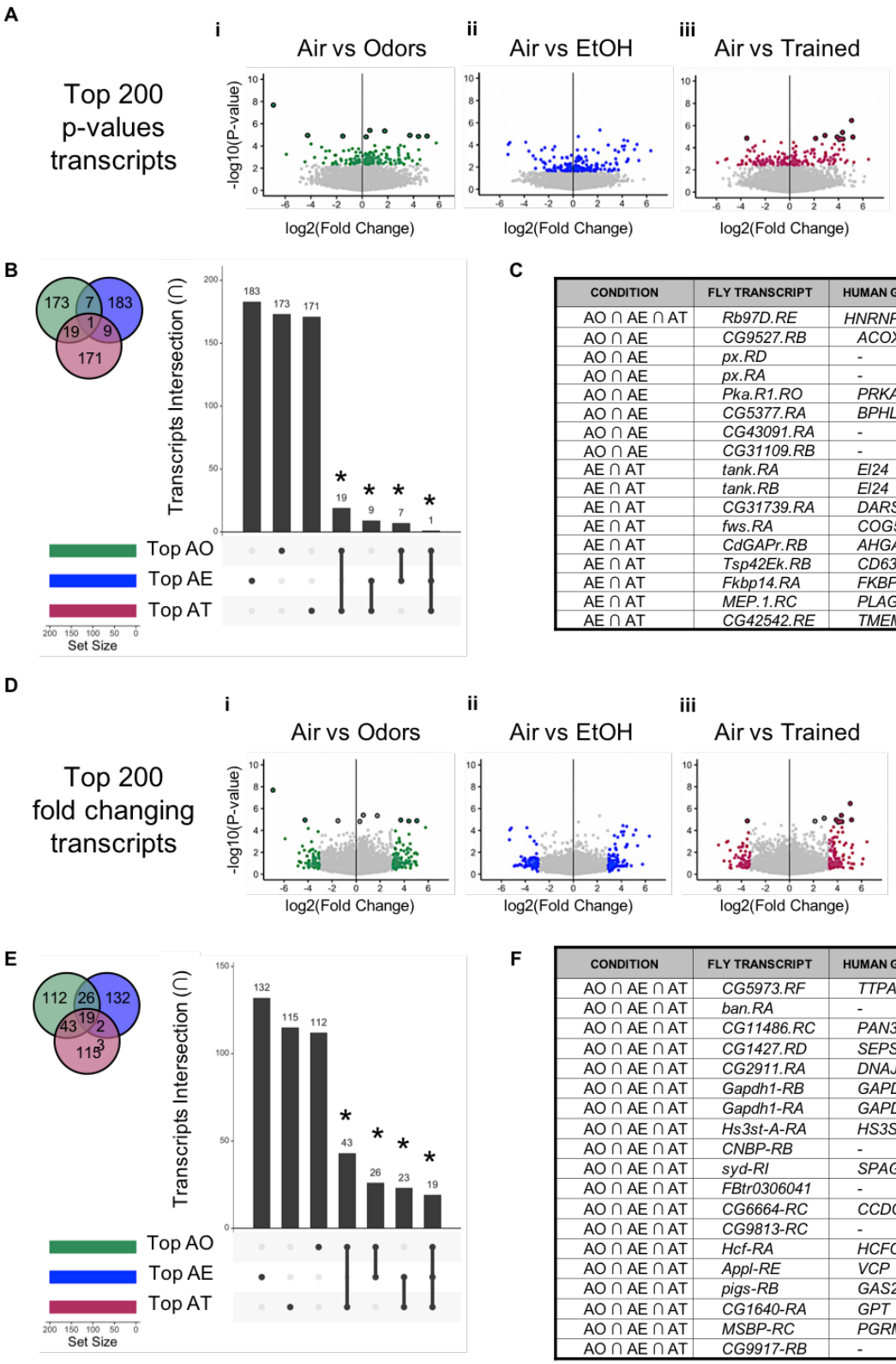
Candidate transcripts underlying ethanol cue memories

This study was motivated by our interest in identifying whether lasting gene expression changes are associated with formation of ethanol cue memories. As our goal was to investigate how ethanol affected gene expression in memory circuits during memory formation, we were most intrigued by the comparisons between Odor vs Ethanol, Ethanol vs Trained, and Odor vs Trained flies.

We found seven transcripts (six genes) that were differentially expressed in response to Ethanol treatment as compared to Odors only controls (Figure 5A). One transcript was downregulated – *CG17982*$^{-RA}$ – and six were upregulated – translational regulator *Dodeca-satellite-binding protein 1* (*Dp1*$^{-RH}$), transcription factor *elbow B* (*elB*$^{-RD}$), transcription factor *modifier of mdge 4* (*pre-mod(mdg4)-G*$^{-RA}$), hypoxia-induced protein *CG11825*$^{-RA}$, and genes of unknown function *CG17982*$^{-RB}$, and *CG17982*$^{-RA}$.

We found four transcripts (three genes, all of unknown function) that were differentially expressed in response to Trained treatment as compared to Ethanol alone (Figure 5B). One transcript was downregulated – *CG10809*$^{-RB}$ – and three were upregulated – *CG11970*$^{-RA}$, *CG13055*$^{-RB}$, *CG10809*$^{-RA}$.

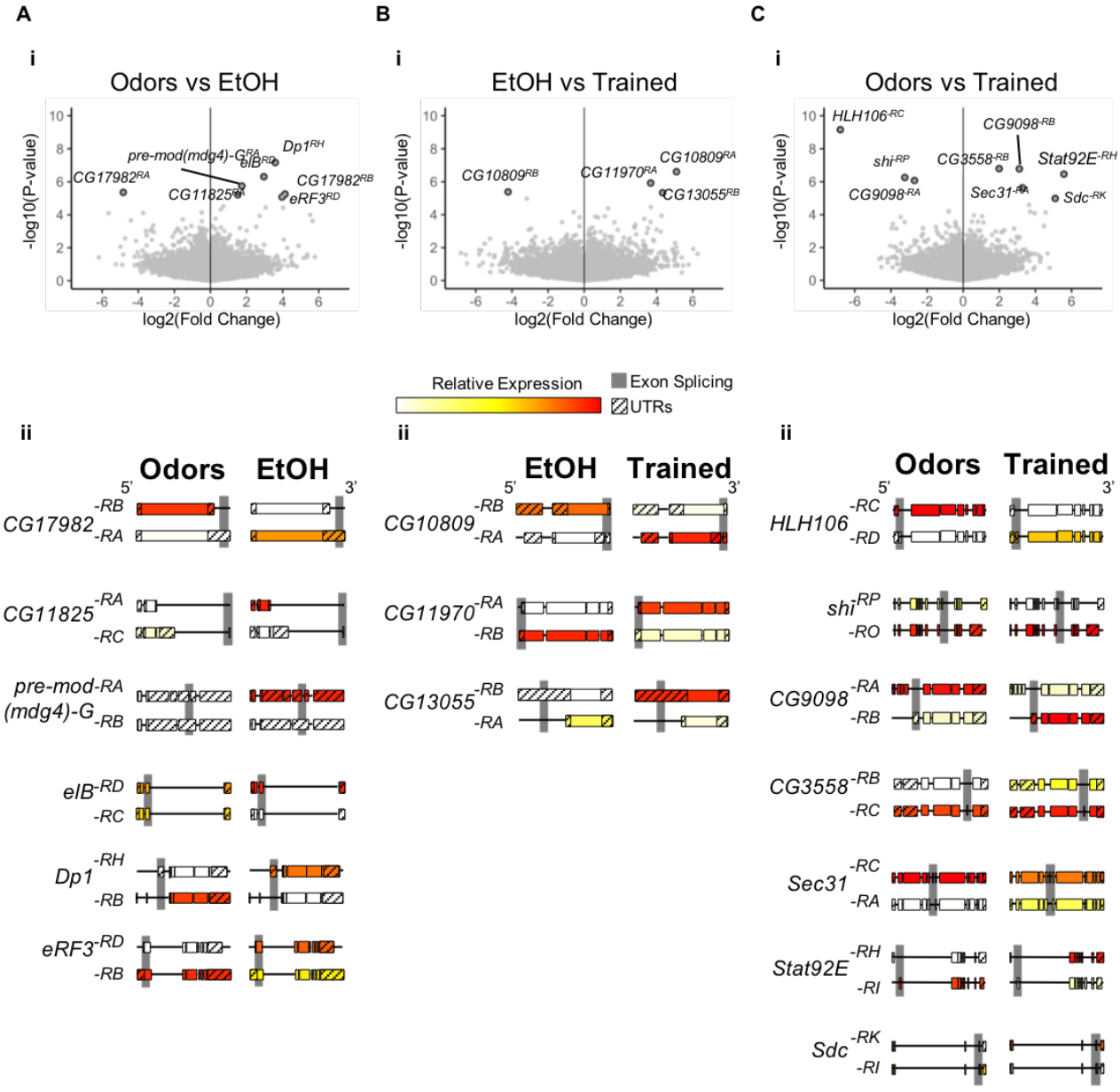
We found eight transcripts (seven genes) that were differentially expressed in response to Trained treatment as compared to Odors only controls (Figure 5C, Supplemental Figure 3). Three of these transcripts were downregulated – the Sterol Regulatory Element Binding Protein SREBP (*HLH106*$^{-RC}$), the *Drosophila* homolog of dynamin *shibire* (*shi*$^{-RP}$), and a predicted Ras guanine-nucleotide exchange factor *CG9098*$^{-RA}$. Four transcripts were upregulated – a retinoic acid-like protein *CG3558*$^{-RB}$, the same predicted Ras guanine-nucleotide exchange factor *CG9098*$^{-RB}$, a Coat Protein Complex II factor *Secretory 31* (*Sec31*$^{-RA}$), a transmembrane heparin sulfate proteoglycan *Syndecan* (*Sdc*$^{-RK}$), and a transcription factor that functions in the JAK/STAT pathway *Signal-*



CONDITION	FLY TRANSCRIPT	HUMAN GENE
AO ∩ AE ∩ AT	*Rb97D.RE*	*HNRNPA2B1*
AO ∩ AE	*CG9527.RB*	*ACOX3*
AO ∩ AE	*px.RD*	-
AO ∩ AE	*px.RA*	-
AO ∩ AE	*Pka.R1.RO*	*PRKAR1A*
AO ∩ AE	*CG5377.RA*	*BPHL*
AO ∩ AE	*CG43091.RA*	-
AO ∩ AE	*CG31109.RB*	-
AE ∩ AT	*tank.RA*	*EI24*
AE ∩ AT	*tank.RB*	*EI24*
AE ∩ AT	*CG31739.RA*	*DARS2*
AE ∩ AT	*fws.RA*	*COG5*
AE ∩ AT	*CdGAPr.RB*	*AHGAP32*
AE ∩ AT	*Tsp42Ek.RB*	*CD63*
AE ∩ AT	*Fkbp14.RA*	*FKBP14*
AE ∩ AT	*MEP.1.RC*	*PLAGL2*
AE ∩ AT	*CG42542.RE*	*TMEM165*

CONDITION	FLY TRANSCRIPT	HUMAN GENE
AO ∩ AE ∩ AT	*CG5973.RF*	*TTPAL*
AO ∩ AE ∩ AT	*ban.RA*	-
AO ∩ AE ∩ AT	*CG11486.RC*	*PAN3*
AO ∩ AE ∩ AT	*CG1427.RD*	*SEPSECS*
AO ∩ AE ∩ AT	*CG2911.RA*	*DNAJC24*
AO ∩ AE ∩ AT	*Gapdh1-RB*	*GAPDH*
AO ∩ AE ∩ AT	*Gapdh1-RA*	*GAPDH*
AO ∩ AE ∩ AT	*Hs3st-A-RA*	*HS3ST5*
AO ∩ AE ∩ AT	*CNBP-RB*	-
AO ∩ AE ∩ AT	*syd-RI*	*SPAG9*
AO ∩ AE ∩ AT	*FBtr0306041*	-
AO ∩ AE ∩ AT	*CG6664-RC*	*CCDC6*
AO ∩ AE ∩ AT	*CG9813-RC*	-
AO ∩ AE ∩ AT	*Hcf-RA*	*HCFC1*
AO ∩ AE ∩ AT	*Appl-RE*	*VCP*
AO ∩ AE ∩ AT	*pigs-RB*	*GAS2L1*
AO ∩ AE ∩ AT	*CG1640-RA*	*GPT*
AO ∩ AE ∩ AT	*MSBP-RC*	*PGRMC1*
AO ∩ AE ∩ AT	*CG9917-RB*	-

Figure 4. Intersectional analysis of top differentially expressed transcripts across treatments, as compared to Air Controls.
A) Volcano plots of **(i)** Air vs Odors, **(ii)** Air vs EtOH and **(iii)** Air vs Trained comparisons (dark outline, FDR < 0.05). Colors depict the top 200 p-value transcripts. **B)** An upgraded Venn Diagram plot generated by an R package called 'UpSetR' demonstrating intersection in top 200 p-value transcripts (abbreviated by first letter in treatment). **C)** A few of the transcripts from intersectional analysis in **(B)** are listed in a table along with corresponding high DIOPT scoring human genes. **D)** Volcano plots of **(i)** Air vs Odors, **(ii)** Air vs EtOH and **(iii)** Air vs Trained comparisons. Colors depict the top 200 fold changing transcripts. **E)** An UpSetR plot demonstrating intersection in top 200 fold-change differentially expressed transcripts. **F)** A few of the transcripts from intersectional analysis in **(D)** are listed in a table along with corresponding high DIOPT scoring human genes.



FLY GENE	HUMAN GENE	SIMILARITY
CG17982	--	--
eRF3	*GSPT1*	70%
Dp1	*HDLBP*	67%
elB	*ZNF703*	35%
pre-mod(mdg4)-G	*FLYWCH1*	44%
CG11825	*HIGD1C*	65%
CG17982	--	--
eRF3	*GSPT1*	70%

FLY GENE	HUMAN GENE	SIMILARITY
CG10809	*ANKRD54*	44%
CG11970	*NFRKB*	40%
CG13055	--	--

FLY GENE	HUMAN GENE	SIMILARITY
HLH106	*SREBF2*	47%
shi	*DNM1*	78%
CG9098	*SH2D3C*	45%
CG3558	*FAM160A2*	42%
Sec31	*SEC31A*	51%
Stat92E	*STAT5B*	48%
Sdc	*SDC2*	41%

Figure 5. Differential transcripts across particular pairwise treatment comparisons.
(i) Volcano plots of **(A)** Odors vs EtOH, **(B)** EtOH vs Trained and **(C)** Odors vs Trained comparisons (dark outline, FDR < 0.05). **(ii)** Significantly altered transcripts are depicted alongside another highly expressed isoform of the same gene, with gray shaded regions denoting splicing differences between isoforms. **(iii)** Transcripts are listed in a table along with corresponding high DIOPT scoring human genes.

transducer and activator of transcription protein at 92E (*Stat92E-RH*). The predicted splice junctions associated with these transcripts suggest a complex pattern of expression (Supplemental Figure 6). Using the DRSC Integrative Ortholog Prediction Tool (DIOPT) (Hu *et al.* 2011) we identified, when possible, the human genes corresponding to the implicated fly genes (Figure 5Aiii,Biii,Ciii).

Molecular interactions with candidate ethanol cue memory genes

To better visualize and probe deeper into the known functions of differentially expressed transcripts associated with formation of ethanol cue memories, we used the open source Cytoscape platform (Shannon *et al.* 2003) and MIST database (Hu *et al.* 2018) to identify proteins that associate or interact with Odor vs Trained genes of interest (Figure 6). By overlaying our dataset information like expression level and fold change, onto this network, we provide a foundation for prospective hypothesis-driven bioinformatic inquiries and experiments. Notably, Stat92E had the largest known network, including a number of proteins associated with nucleosome remodeling (HDAC1, brm, mor), transcriptional regulation (Taf1, Ada2b, kay, Jra, CG13510), and splicing (Cdc5, CG7564, tsu,). Altogether the novel targets, and their interactors, offer experimental avenues to explore in future *Drosophila* and mammalian studies.

Discussion

The nuclear and cell-type specific nature of our data provide a unique analysis of genes actively transcribed in the nucleus of memory-encoding neurons 24 hours after experience with alcohol or memory formation. Selection of poly-adenylated transcripts further allowed us to investigate changes to processed transcripts, rather than nascent RNA.

Known genes implicated in alcohol experience and memory formation.

Two of the genes that express alternative transcript isoforms in our ethanol cue memory data have been functionally implicated in similar behaviors. *shi,* the fly homolog of Dynamin (van der Bliek and Meyerowitz 1991) plays a role in endocytosis critical for synaptic transmission in MB neurons (McGuire *et al.* 2001; Kasuya *et al.* 2009) and in ethanol tolerance (Krishnan *et al.* 2012). We've also previously shown that expressing a dominant negative form of *shi* in MB neurons disrupts ethanol cue memory (Kaun et al, 2011). Stat92E, a transcription factor in the JAK/STAT signaling pathway, plays a role in long-term memory in *Drosophila* (Copf et al. 2011). We've recently shown that decreasing Stat92E in MB neurons similarly reduces ethanol cue memory (Petruccelli *et al.* 2018). To our knowledge, very few gene targets isolated from this study have been implicated in long-term

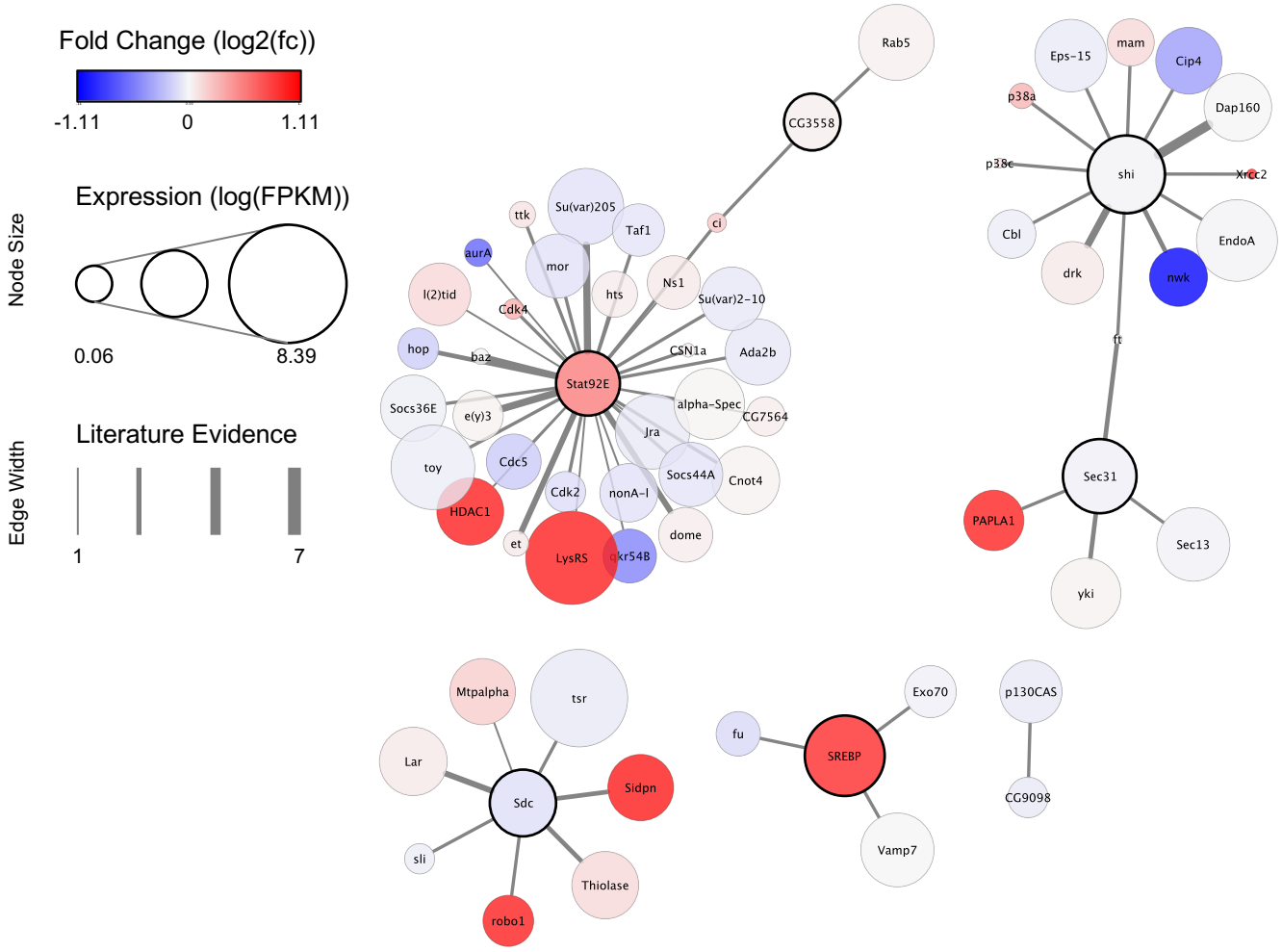


Figure 6. Protein interaction network of proteins associated with Odor vs Trained significant transcripts.
Each circle node represents a protein, with dark outlines denoting proteins associated with significant Odor vs Trained transcripts. Attributes of the nodes includes fold change as color, expression level as size and edge thickness between nodes represents the extent of protein-protein interaction evidence from the MIST database.

memory in *Drosophila* (Supplemental Figure 7). However, a recent study showing the effects of acute response to alcohol shows some overlap in transcriptomic targets (Supplemental Figure 7) (SIGNOR AND NUZHDIN 2018).

Context specific transcript expression
An intriguing pattern that emerged from our data is that administration of repeated mild doses of alcohol, odors alone, or odors paired with ethanol all produced significantly altered transcript isoforms. Remarkably, the particular isoforms identified were specific to the behavioral paradigm the animals experienced. Even altered transcripts associated with the presentation of alcohol alone were very different from those when alcohol was presented concomitantly with an odor. This suggests that transcript isoform expression is specific to the type of memory formed.

Experience-dependent alternative splicing in MB neurons
Alternative transcript isoforms can be expressed through alternative transcription start sites or alternative splicing. Intriguingly, there are known protein-protein interactions between Stat92E and spliceosome complex associated proteins, including Cdc5, CG7564, and tsu (GURUHARSHA *et al.* 2011) (Figure 6). This suggests that splicing may readily occur in MB neurons as a result of experience. A caveat to our data, however, is that because we did not *a priori* expect to see splicing differences, our sequencing data was not performed with long-read sequencing techniques. Nor were statistical considerations, such as k-means clustering of transcripts or weighting exon-exon junction reads, used to compare splicing across genes.

The most direct cause of abnormal splicing is due to a mutation in the core splicing consensus sequences (CIEPLY AND CARSTENS 2015). Our data, however, are derived from animals with the same genetic backgrounds and rearing, and demonstrate very few altered transcripts are detected 24 hours after experience. This suggests that mutagenesis is not the cause of the potential splicing differences we observed. Furthermore, although our data showed no direct evidence that spliceosomal machinery was differentially expressed with treatment, this does not rule out the possibility that spliceosome-associated proteins were acutely affected during treatment exposure. Our data may represent the aftermath of these, or post-translational response, effects.

Splicing associated with memory formation
Splicing of particular targets in response to memory formation is not without precedent. For example, the splicing of the *Orb2A* transcript is required for both associative appetitive and courtship suppression memory in *Drosophila* (GILL *et al.* 2017). There is also evidence that this phenomenon may be conserved across species, since contextual fear conditioning induces differential splicing in the hippocampus of mice (POPLAWSKI *et al.* 2016). It is, therefore, conceivable that whereas splicing may occur during all forms of synaptic plasticity, the isoform targets may be specific to the type of memory formed. The broad implication of this prediction is that the molecular engram of memory within relevant cells uses transcriptional diversity to provide the cellular plasticity associated with the type of memory being formed. This cellular encoding is bolstered by the diversify of epigenetic changes caused by neuronal activity associated with memory formation. For example, activity-induced histone modifications caused a late-onset shift in Neurexin-1 splicing reduced the stability of memories (DING *et al.* 2017). Our data suggest that alcohol can affect transcriptional events, and thus shape how context is encoded during formation of memories.

Alcohol regulates splicing in cue-encoding neurons
Several recent studies have shown that alternative splicing is correlated to chronic alcohol consumption. Chronic self-administration of alcohol in cynomolgus macaques (*Macaca fascicularis*) is associated with alternative splicing of AMPA subunits in the prefrontal cortex (ACOSTA *et al.* 2011). Broader transcriptomic analysis demonstrated over-representation of genes associated with splicing across brain regions in chronic ethanol self-administrating rhesus macaques (*Macaca mulatta*) (IANCU *et al.* 2018). Similar alcohol-induced effects on splicing extend to humans (FARRIS AND MAYFIELD 2014). Post-mortem analysis of the brains of AUD patients showed novel splicing in *GABAB1* that decreased expression of the GABA binding site (LEE *et al.* 2014). Similarly, alcohol-induced splicing events also occurred in the developing human cortex *in utero*, potentially resulting in devastating neurodevelopmental consequences (KAWASAWA *et al.* 2017).

Our data suggest that repeated alcohol exposures have lasting consequences on the transcript isoforms expressed in *Drosophila*. Importantly, altered isoform transcription occurred in a diversity of genes, including transcription factors like Stat92E, which could have long-lasting and broad effects on transcription within memory circuits (COPF *et al.* 2011). As alcohol alone did not induce expression of alternative transcripts of Stat92E, this suggests that splicing of this gene might occur due to memory formation. Since our analysis was restricted to neurons necessary for memories associated with punishment or reward, it is possible that even small effects could have widespread consequences for subsequent memory formation and vulnerability to dependence on drugs of abuse. Uncovering whether this is a broader phenomenon will require identifying lasting transcriptional events that occur in the same cell types across a diversity of memory and drug exposure paradigms.

Diverse splicing mechanisms across addictions
Alternative splicing has been recently implicated in cocaine addiction (CATES *et al.* 2018). Transcription factor E2F3a regulates cocaine-induced alternative splicing in the mouse nucleus accumbens, and E2F3b mediates cocaine responses in the prefrontal cortex (CATES *et al.* 2019). *Drosophila E2F1* was expressed in our MB nuclei dataset (Figure 2cii), although it was not alternatively spliced in response to alcohol or alcohol-cue training (See transcript count data in GEO). We speculate that cocaine utilizes the E2F transcription factor family more than alcohol does to drive drug-cue memory formation and retention. Future investigations will likely identify both conserved drug-specific and convergent molecular mechanisms that influence transcriptional activity in reward circuitry.

End Materials

Data Availability
Raw sequencing data and count data presented in this work will be made freely available on NCBI Gene Expression Omnibus via accession number GSE108525.

Author Contributions and Notes

E.P. and K.R.K. designed the research and wrote the manuscript. E.P. and N.L. performed INTACT nuclear extraction. N.L. developed the pipeline to align RNA-seq reads. E.P. performed statistical analyses and produced figures in consultation with K.R.K. Revision of the manuscript was performed by E.P., N.L., and K.R.K.

The authors declare no conflict of interest.

Supporting information is addended following references of this article.

Acknowledgments

We thank Lee Henry (CSHL) for advice on INTACT, and members of the Kaun Lab, especially Reza Azanchi, Michael Feyder, and Yanabah Jaques for assistance with *Drosophila* husbandry and lab tasks. We thank Dr. Faith Liebl (SIUE), Dr. Kate O'Connor-Giles (Brown University), and members of the Kaun lab for helpful feedback on the manuscript. This work was supported by the Richard and Susan Smith Family Foundation, Newton, MA, the Carney Institute for Brain Science Center of Biomedical Research Excellence "Center for Nervous System Function" (NIGMS P20GM103645), and the National Institute on Alcohol Abuse and Alcoholism (R01AA024434).

References

Acosta, G., D. P. Freidman, K. A. Grant and S. E. Hemby, 2011 Alternative splicing of AMPA subunits in prefrontal cortical fields of cynomolgus monkeys following chronic ethanol self-administration. Front Psychiatry 2**:** 72.

Alberini, C. M., and E. R. Kandel, 2014 The regulation of transcription in memory consolidation. Cold Spring Harb Perspect Biol 7**:** a021741.

Aso, Y., D. Hattori, Y. Yu, R. M. Johnston, N. A. Iyer *et al.*, 2014 The neuronal architecture of the mushroom body provides a logic for associative learning. Elife 3**:** e04577.

Bailey, C. H., D. Bartsch and E. R. Kandel, 1996 Toward a molecular definition of long-term memory storage. Proc Natl Acad Sci U S A 93**:** 13445-13452.

Biever, A., P. G. Donlin-Asp and E. M. Schuman, 2019 Local translation in neuronal processes. Curr Opin Neurobiol 57**:** 141-148.

Bolger, A. M., M. Lohse and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30**:** 2114-2120.

Bramham, C. R., and D. G. Wells, 2007 Dendritic mRNA: transport, translation and function. Nat Rev Neurosci 8**:** 776-789.

Cates, H. M., R. C. Bagot, E. A. Heller, I. Purushothaman, C. K. Lardner *et al.*, 2019 A novel role for E2F3b in regulating cocaine action in the prefrontal cortex. Neuropsychopharmacology 44**:** 776-784.

Cates, H. M., E. A. Heller, C. K. Lardner, I. Purushothaman, C. J. Pena *et al.*, 2018 Transcription Factor E2F3a in Nucleus Accumbens Affects Cocaine Action via Transcription and Alternative Splicing. Biol Psychiatry 84**:** 167-179.

Cheng, H. G., H. Kaakarli, J. Breslau and J. C. Anthony, 2018 Assessing Changes in Alcohol Use and Alcohol Use Disorder Prevalence in the United States: Evidence From National Surveys From 2002 Through 2014. JAMA Psychiatry 75**:** 211-213.

Cieply, B., and R. P. Carstens, 2015 Functional roles of alternative splicing factors in human disease. Wiley Interdiscip Rev RNA 6**:** 311-326.

Clemens, K. J., and N. M. Holmes, 2018 An extended history of drug self-administration results in multiple sources of control over drug seeking behavior. Prog Neuropsychopharmacol Biol Psychiatry 87**:** 48-55.

Copf, T., V. Goguel, A. Lampin-Saint-Amaux, N. Scaplehorn and T. Preat, 2011 Cytokine signaling through the JAK/STAT pathway is required for long-term memory in Drosophila. Proc Natl Acad Sci U S A 108**:** 8059-8064.

Courtney, K. E., J. P. Schacht, K. Hutchison, D. J. Roche and L. A. Ray, 2016 Neural substrates of cue reactivity: association with treatment outcomes and relapse. Addict Biol 21**:** 3-22.

Crocker, A., X. J. Guan, C. T. Murphy and M. Murthy, 2016 Cell-Type-Specific Transcriptome Analysis in the Drosophila Mushroom Body Reveals Memory-Related Changes in Gene Expression. Cell Rep 15**:** 1580-1596.

Devineni, A. V., and U. Heberlein, 2009 Preferential ethanol consumption in Drosophila models features of addiction. Curr Biol 19**:** 2126-2132.

Ding, X., S. Liu, M. Tian, W. Zhang, T. Zhu *et al.*, 2017 Activity-induced histone modifications govern Neurexin-1 mRNA splicing and memory preservation. Nat Neurosci 20**:** 690-699.

Erickson, E. K., E. K. Grantham, A. S. Warden and R. A. Harris, 2019 Neuroimmune signaling in alcohol use disorder. Pharmacol Biochem Behav 177**:** 34-60.

Farris, S. P., and R. D. Mayfield, 2014 RNA-Seq reveals novel transcriptional reorganization in human alcoholic brain. Int Rev Neurobiol 116**:** 275-300.

Frazee, A. C., G. Pertea, A. E. Jaffe, B. Langmead, S. L. Salzberg *et al.*, 2015 Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat Biotechnol 33**:** 243-246.

Gill, J., Y. Park, J. P. McGinnis, C. Perez-Sanchez, M. Blanchette *et al.*, 2017 Regulated Intron Removal Integrates Motivational State and Experience. Cell 169**:** 836-848 e815.

Grant, B. F., S. P. Chou, T. D. Saha, R. P. Pickering, B. T. Kerridge *et al.*, 2017 Prevalence of 12-Month Alcohol Use, High-Risk Drinking, and DSM-IV Alcohol Use Disorder in the United States, 2001-2002 to 2012-2013: Results From the National Epidemiologic Survey on Alcohol and Related Conditions. JAMA Psychiatry 74**:** 911-923.

Groefsema, M., R. Engels and M. Luijten, 2016 The role of social stimuli content in neuroimaging studies investigating alcohol cue-reactivity. Addict Behav 58**:** 123-128.

Guruharsha, K. G., J. F. Rual, B. Zhai, J. Mintseris, P. Vaidya *et al.*, 2011 A protein complex network of Drosophila melanogaster. Cell 147**:** 690-703.

Henry, G. L., F. P. Davis, S. Picard and S. R. Eddy, 2012 Cell type-specific genomics of Drosophila neurons. Nucleic Acids Res 40**:** 9691-9704.

Hu, Y., I. Flockhart, A. Vinayagam, C. Bergwitz, B. Berger *et al.*, 2011 An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics 12**:** 357.

Hu, Y., A. Vinayagam, A. Nand, A. Comjean, V. Chung *et al.*, 2018 Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data. Nucleic Acids Res 46**:** D567-D574.

Iancu, O. D., A. Colville, N. A. R. Walter, P. Darakjian, D. L. Oberbeck *et al.*, 2018 On the relationships in rhesus macaques between chronic ethanol consumption and the brain transcriptome. Addict Biol 23**:** 196-205.

Jasinska, A. J., E. A. Stein, J. Kaiser, M. J. Naumer and Y. Yalachkov, 2014 Factors modulating neural reactivity to drug

cues in addiction: a survey of human neuroimaging studies. Neurosci Biobehav Rev 38: 1-16.
Kasuya, J., H. Ishimoto and T. Kitamoto, 2009 Neuronal mechanisms of learning and memory revealed by spatial and temporal suppression of neurotransmission using shibire, a temperature-sensitive dynamin mutant gene in Drosophila melanogaster. Front Mol Neurosci 2: 11.
Kaun, K. R., R. Azanchi, Z. Maung, J. Hirsh and U. Heberlein, 2011 A Drosophila model for alcohol reward. Nat Neurosci 14: 612-619.
Kawasawa, Y. I., S. Mohammad, A. I. Son, H. Morizono, A. Basha *et al.*, 2017 Genome-wide profiling of differentially spliced mRNAs in human fetal cortical tissue exposed to alcohol. Alcohol 62: 1-9.
Kendler, K. S., H. Ohlsson, J. Sundquist and K. Sundquist, 2016 Alcohol Use Disorder and Mortality Across the Lifespan: A Longitudinal Cohort and Co-relative Analysis. JAMA Psychiatry 73: 575-581.
Krishnan, H. R., Y. M. Al-Hasan, J. B. Pohl, A. Ghezzi and N. S. Atkinson, 2012 A role for dynamin in triggering ethanol tolerance. Alcohol Clin Exp Res 36: 24-34.
Lacar, B., S. B. Linker, B. N. Jaeger, S. R. Krishnaswami, J. J. Barron *et al.*, 2016 Nuclear RNA-seq of single neurons reveals molecular signatures of activation. Nat Commun 7: 11022.
Lee, C., R. D. Mayfield and R. A. Harris, 2014 Altered gamma-aminobutyric acid type B receptor subunit 1 splicing in alcoholics. Biol Psychiatry 75: 765-773.
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
Logge, W. B., K. C. Morley, P. S. Haber and A. J. Baillie, 2019 Executive Functioning Moderates Responses to Appetitive Cues: A Study in Severe Alcohol Use Disorder and Alcoholic Liver Disease. Alcohol Alcohol 54: 38-46.
McGuire, S. E., P. T. Le and R. L. Davis, 2001 The role of Drosophila mushroom body signaling in olfactory memory. Science 293: 1330-1333.
Nakahata, Y., and R. Yasuda, 2018 Plasticity of Spine Structure: Local Signaling, Translation and Cytoskeletal Reorganization. Front Synaptic Neurosci 10: 29.
Nestler, E. J., 2013 Cellular basis of memory for addiction. Dialogues Clin Neurosci 15: 431-443.
Nunez, K. M., R. Azanchi and K. R. Kaun, 2018 Cue-Induced Ethanol Seeking in Drosophila melanogaster Is Dose-Dependent. Front Physiol 9: 438.
Pan, Q., O. Shai, L. J. Lee, B. J. Frey and B. J. Blencowe, 2008 Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40: 1413-1415.
Pankova, K., and A. Borst, 2016 RNA-Seq Transcriptome Analysis of Direction-Selective T4/T5 Neurons in Drosophila. PLoS One 11: e0163986.
Pertea, M., D. Kim, G. M. Pertea, J. T. Leek and S. L. Salzberg, 2016 Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11: 1650-1667.
Petruccelli, E., M. Feyder, N. Ledru, Y. Jaques, E. Anderson *et al.*, 2018 Alcohol Activates Scabrous-Notch to Influence Associated Memories. Neuron 100: 1209-1223 e1204.
Poplawski, S. G., L. Peixoto, G. S. Porcari, M. E. Wimmer, A. G. McNally *et al.*, 2016 Contextual fear conditioning induces differential alternative splicing. Neurobiol Learn Mem 134 Pt B: 221-235.
Richter, J. D., 2010 Translational control of synaptic plasticity. Biochem Soc Trans 38: 1527-1530.
Ron, D., and S. Barak, 2016 Molecular mechanisms underlying alcohol-drinking behaviours. Nat Rev Neurosci 17: 576-591.
Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang *et al.*, 2003 Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504.
Shigeoka, T., H. Jung, J. Jung, B. Turner-Bridger, J. Ohk *et al.*, 2016 Dynamic Axonal Translation in Developing and Mature Visual Circuits. Cell 166: 181-192.
Shih, M. M., F. P. Davis, G. L. Henry and J. Dubnau, 2019 Nuclear Transcriptomes of the Seven Neuronal Cell Types That Constitute the Drosophila Mushroom Bodies. G3 (Bethesda) 9: 81-94.
Signor, S., and S. Nuzhdin, 2018 Dynamic changes in gene expression and alternative splicing mediate the response to acute alcohol exposure in Drosophila melanogaster. Heredity (Edinb) 121: 342-360.
Sweatt, J. D., 2016 Neural plasticity and behavior - sixty years of conceptual advances. J Neurochem 139 Suppl 2: 179-199.
Trudell, J. R., R. O. Messing, J. Mayfield and R. A. Harris, 2014 Alcohol dependence: molecular and behavioral evidence. Trends Pharmacol Sci 35: 317-323.
Uchida, S., and G. P. Shumyatsky, 2018 Synaptically Localized Transcriptional Regulators in Memory Formation. Neuroscience 370: 4-13.
Valyear, M. D., F. R. Villaruel and N. Chaudhri, 2017 Alcohol-seeking and relapse: A focus on incentive salience and contextual conditioning. Behav Processes 141: 26-32.
van der Bliek, A. M., and E. M. Meyerowitz, 1991 Dynamin-like protein encoded by the Drosophila shibire gene associated with vesicular traffic. Nature 351: 411-414.
Wickham, H., 2009 *Ggplot2 : elegant graphics for data analysis*. Springer, New York.
Widmer, Y. F., A. Bilican, R. Bruggmann and S. G. Sprecher, 2018 Regulators of Long-Term Memory Revealed by Mushroom Body-Specific Gene Expression Profiling in Drosophila melanogaster. Genetics 209: 1167-1181.
Wolf, F. W., A. R. Rodan, L. T. Tsai and U. Heberlein, 2002 High-resolution analysis of ethanol-induced locomotor stimulation in Drosophila. J Neurosci 22: 11035-11044.

Ethanol causes lasting differential transcription in mushroom body neurons

Supplemental Information

Authors: Emily Petruccelli*†, Nicolas Ledru*, Karla R. Kaun*

Authors' Affiliations: *Department of Neuroscience, Brown University, Providence, Rhode Island, USA, 02912. †Current address: Department of Biological Sciences, Southern Illinois University Edwardsville, Edwardsville, Illinois, USA, 62026.

Table of Contents **page**

Supplemental Title page….....1

Supplemental Figure 1.....….2
INTACT/MB010B is expressed in MB neurons and does not disrupt alcohol memory

Supplemental Figure 2.....….3
Table summarizing QualiMap (BAM QC) analysis of the RNA-seq libraries

Supplemental Figure 3.....….4
Table summarizing hypergeometric statistics demonstrating overlap in expression across treatments

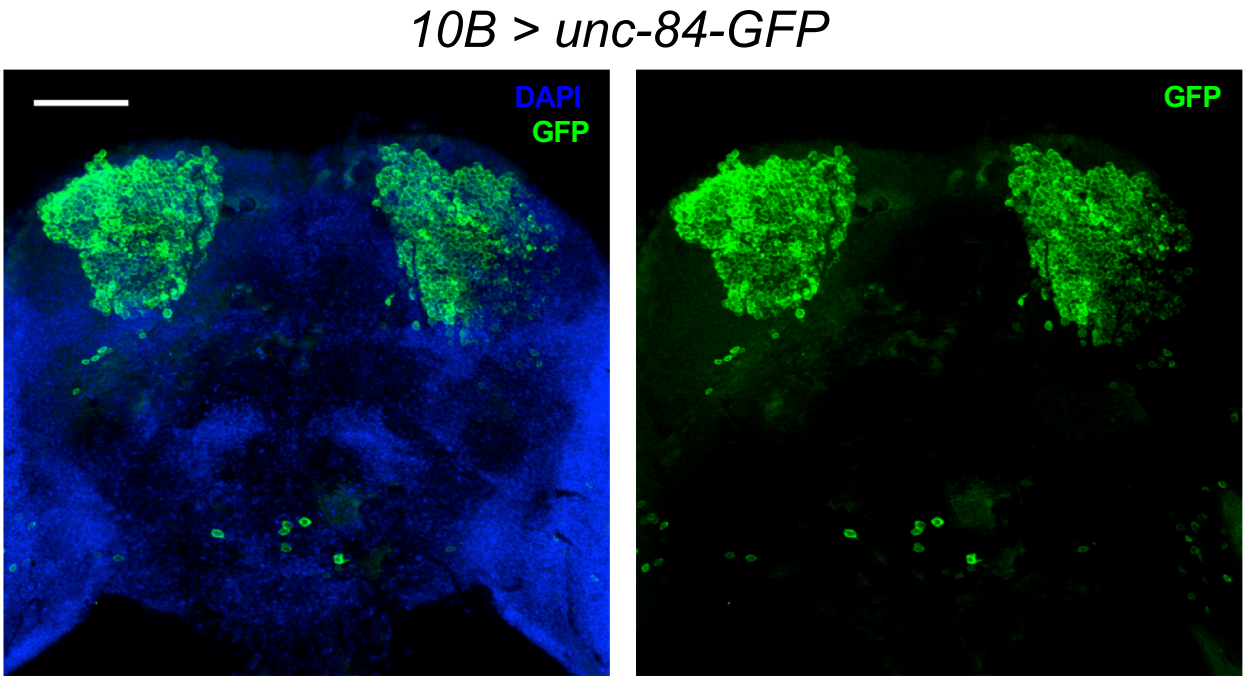
Supplemental Figure 4.....….5
Intersectional analysis of top differentially expressed transcripts across treatments

Supplemental Figure 5.....6
Table summarizing DAVID analysis of gene ontology (GO) of top 200 p-value genes

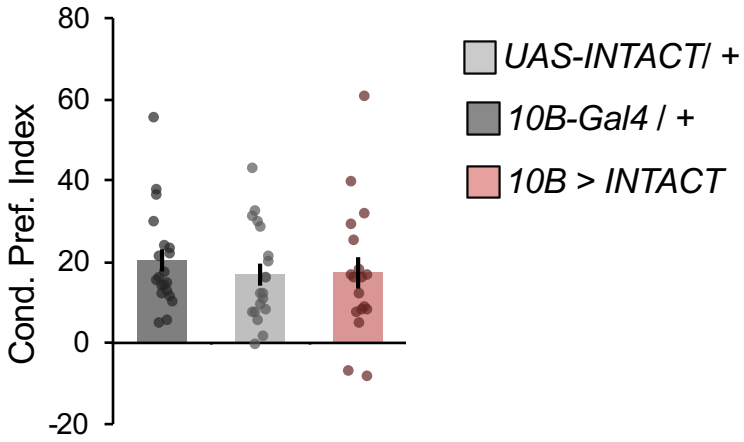
Supplemental Figure 6….....7
Sashimi plots for transcripts differentially expressed between Trained and Odor treatments

Supplemental Figure 7….....8
Table comparing our results to recent, similar RNA-Seq studies.

A



B



Supplemental Figure 1. Restricted expression of INTACT transgene to all MB nuclei. A) A representative immunohistochemical staining of a 3-5-day-old adult male brain (DAPI staining DNA, Blue; anti-GFP staining GFP, Green). Confocal max stack image taken at 20x resolution (Scale bar 50 um). **B)** Expression of the INTACT transgene (unc^{84}-GFP) in MB neurons did not interfere with alcohol associative preference; Gal4 control (n = 19), UAS control (n= 20), Gal4>UAS experimental line (n= 18).

	Total Sequences	Uniquely Mapped	Exonic / Intronic / Intergenic	Transcript Counts (FPKM)
Air	33844558	85.87%	97% / 1.5% / 0.7%	15784305
	37211293	87.31%	97% / 1.9% / 0.7%	15262082
	46395515	76.84%	97% / 1.8% / 0.8%	17714446
EtOH	65772590	85.91%	93% / 4.1% / 2.5%	12129006
	25243988	87.39%	97% / 1.6% / 0.9%	14165946
	29887470	83.14%	97% / 1.7% / 0.9%	15497176
Odors	30059906	89.08%	97% / 1.4% / 0.8%	15415177
	25082770	90.57%	97% / 1.4% / 0.8%	14377060
	41694314	89.92%	97% / 1.4% / 0.8%	21702932
	29669861	89.92%	97% / 1.6% / 0.9%	15614864
Trained	27938257	89.86%	97% / 1.4% / 0.8%	16104880
	24496938	89.84%	97% / 1.7% / 1%	14242031
	29179965	90.72%	97% / 1.4% / 0.8%	15455975
	33911395	89.6%	97% / 1.5% / 0.8%	17410191
AVERAGE	**34313487**	**87.5%**		**15776862**

Supplemental Figure 2. QualiMap (BAM QC) analysis of the RNA-seq libraries. The total sequences, uniquely mapped reads %, genomic biotype, and normalized transcript counts (FPKM) are displayed for all 14 libraries.

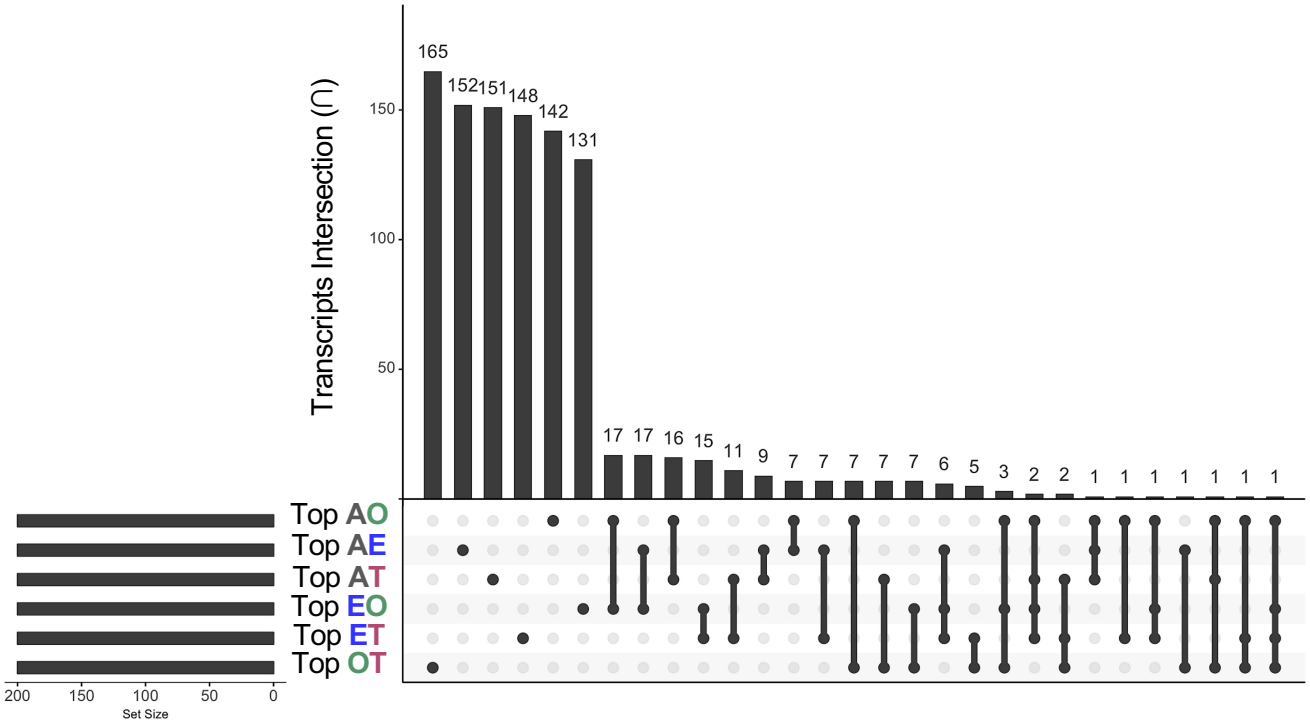
	Top 200 p-Value transcripts	Top 200 Fold Changer transcripts
Interaction	**pValue**	**pValue**
AO ∩ AE	5.6e-4	4.5e-46
AO ∩ AT	2.4e-14	9.3e-73
AE ∩ AT	2.3e-5	9.6e-42

Supplemental Figure 3. Hypergeometric distribution test p-values. The phyper function in R was used to determine hypergeometric p-values for data in text Figure 4B,E. The 'universe' test number was 34,741 transcripts from the current Dm6 transcriptome. Values are as follows: phyper(q, m, n, k, lower.tail=FALSE) … eg. phyper(overlap#, 200, (34741-200), 400, lower.tail=FALSE).

A

	Air vs Odors	Air vs Ethanol	Air vs Trained	Odors vs Ethanol	Ethanol vs Trained	Odors vs Trained
Genes	1	0	0	1	0	0
Transcripts	9	0	10	7	4	8

B

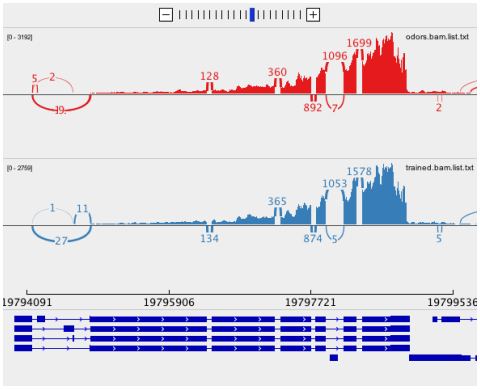


Supplemental Figure 4. Intersectional analysis of top differentially expressed transcripts across treatments, as compared to Air Controls. A) A table representing the number of genes or transcript reaching FDR < 0.05 between all pairwise comparisons: Air vs Odors, Air vs EtOH, Air vs Trained, Odors vs EtOH, EtOH vs Trained, and Odors vs Trained. Colors depict treatment condition. **B)** An upgraded Venn Diagram plot generated by an R package called 'UpSetR' demonstrating the intersection between top 200 p-value transcripts (abbreviated by first letter in treatment).

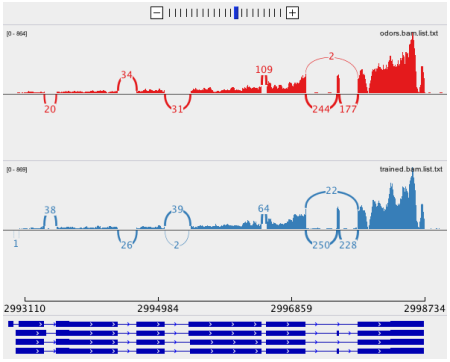
	DAVID GO terms	Adjusted p-value
Air vs Odor	phototransduction	0.86
	NADH dehydrogenase (ubiquinone) activity	0.91
	protein binding	0.92
Air vs Ethanol	**Alternative splicing**	**0.0002**
	Phosphoprotein	**0.041**
	Coiled Coil	**0.047**
Air vs Trained	Alternative splicing	0.062
	Developmental protein	0.2
	Phosphoprotein	0.25

Supplemental Figure 5. DAVID analysis of gene ontology (GO) of top 200 p-value genes. Enrichment of current GO terms within the top 200 p-value genes found between Air vs Odor, Air vs Ethanol, and Air vs Trained pairwise comparisons (Bolded adjusted p-value < 0.05).

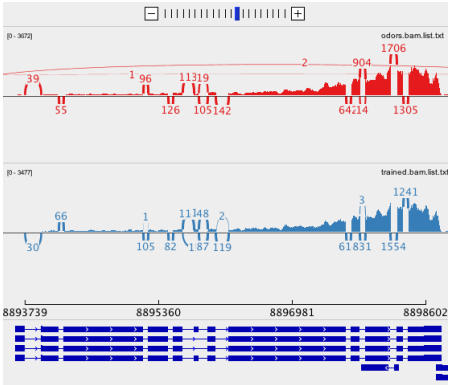
HLH106



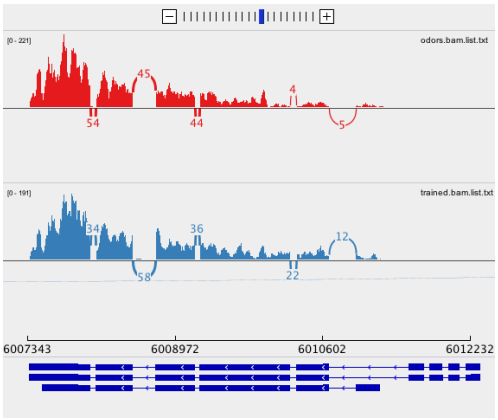
CG3558



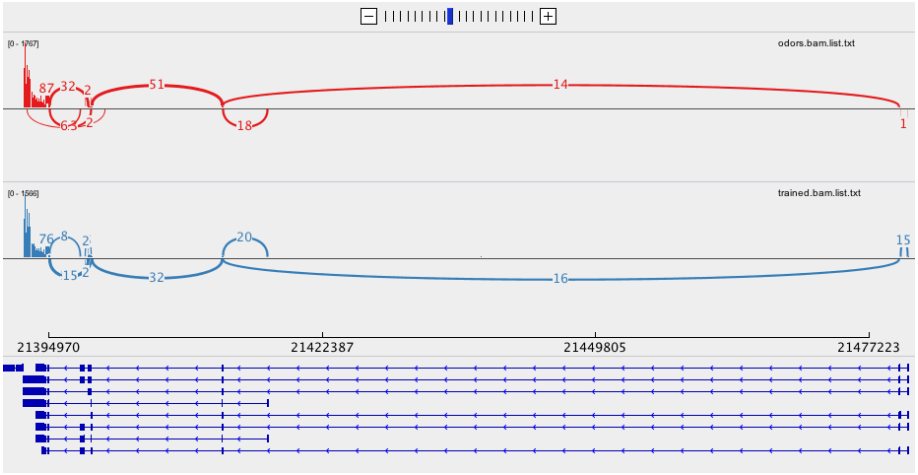
Sec31



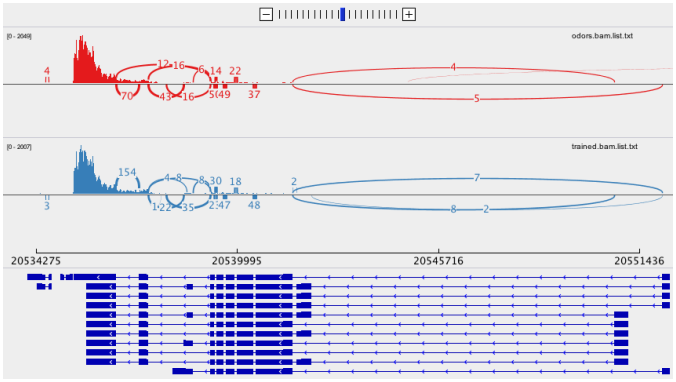
CG9098



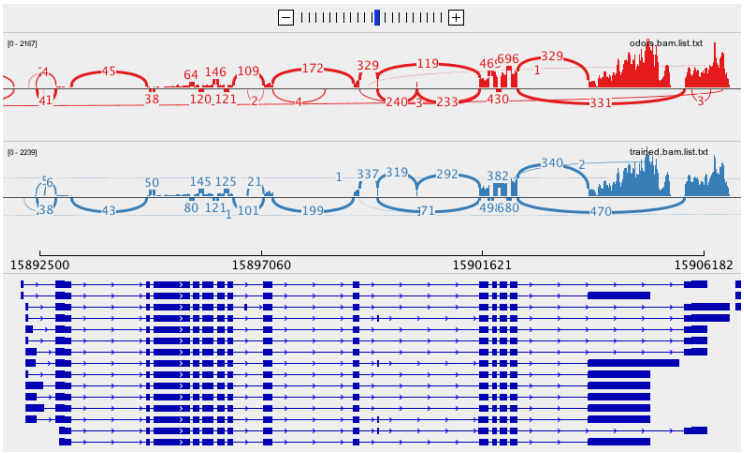
Sdc



Stat92E



shi



Supplemental Figure 6. Sashimi plots from IGV (Broad Institute). Sashimi plots help to visualize splice junctions across merged 'Odors' and 'Trained' samples for differentially expressed transcripts.

Supplemental Figure 7. Comparison of our results to previous RNA-Seq studies.

Publication and study details	Gene lists
Petruccelli et al, 2019 (this study) Time: 24hr after training Tissue: Mushroom body nuclei Condition: appetitive alcohol LTM, 3x 10min (sequential odors) interval spaced 60min Technique: INTACT Analysis: RNA-seq (HISAT2, StringTie, Ballgown)	**Genes associated with top 200 transcripts 'Odors vs Trained':** Aats-trp, aay, Act5C, Adgf-E, Amnionless, AnxB9, asparagine-synthetase, Atg8b, aurA, BBS4, beg, ben, Cad96Cb, CBP, Cdk1, Cdk8, CG10170, CG10428, CG10512, CG10839, CG11131, CG11148, CG11425, CG12116, CG12746, CG12909, CG13484, CG13563, CG13607, CG13708, CG13872, CG14075, CG14275, CG14285, CG14435, CG14621, CG14688, CG14717, CG14879, CG14926, CG15093, CG1513, CG15925, CG1673, CG16756, CG16926, CG16953, CG17570, CG17612, CG17816, CG17821, CG18130, CG2316, CG2911, CG3040, CG30430, CG31419, CG31431, CG3176, CG31787, CG31882, CG32533, CG33098, CG33111, CG33120, CG33795, CG3402, CG34126, CG34166, CG34376, CG34417, CG3505, CG3558, CG42394, CG42502, CG42540, CG42559, CG42577, CG42724, CG42752, CG43093, CG43172, CG43347, CG43630, CG4462, CG4662, CG4842, CG4962, CG5174, CG5446, CG5953, CG5958, CG6067, CG6770, CG8216, CG8508, CG8726, **CG9098**, CG9149, CG9263, CG9300, CG9663, CG9723, CG9777, CG9784, CG9975, CR32205, CR43653, CR43887, CR44265, CR44301, CR44886, CR45113, CR45170, CR45172, CR45884, Crk, crol, cu, Cyp12c1, Cyp4e3, cype, Cyt-b5-r, dbo, Desat2, DhpD, Dip-B, dor, endos, endos, fabp, fit, fus, Glut3, Gpdh, gus, **HLH106**, HmgZ, HPS4, Hr39, Hr4, hubl, Idh, Idh, intr, jnj, l(2)k09913, LKR, mav, mir-4963, mod, mod(mdg4), mRpS18B, mtsh, NaCP60E, net, Not1, nsl1, Obp57b, odd, Or85f, Pask, Pitslre, pix, Rb97D, Reps, Reps, Rm62, Rpl1, RpL3, **Sdc**, **Sec31**, Sh, **shi**, sif, SkpA, snoRNA:Me28S-A2634c, snoRNA:Me28S-C3420b, Sodh-2, **Stat92E**, su(r), Syt1, tau, TFAM, Tgi, Thor, TM4SF, TotA, Trpm, UK114, viaf, Xport ***bold FDR < 0.05**
Crocker et al, 2016 Time: 30min after training Tissue: Kenyon Cells, male only? Condition: aversive shock olfactory LTM, 8x 1min interval spaced 10min Technique: Single Cell RNA-seq Analysis: TopHat2, HTseq-count, DESeq2	no differentially expressed changes found in KCs, likely due to pooling
Widmer et al, 2018 Time: T2 (12-24hr) and T3 (24-48hr) after training Tissue: Heads Condition: appetitive sucrose olfactory LTM, starve 19-21hr then 2min (sequential odors spaced 1min) Technique: Dam-Pol II Analysis: Bowtie2, damidseq_pipeline with FDR via 50,000 simulations	**T2:** Acox57D-p, Acp54A1, Adam, Adk2, Asciz, ATPsynC, Cdk7, CG10344, CG10562, CG10581, CG10584, CG11069, CG11137, CG11597, CG11668, CG11693, CG11777, CG11852, CG12118, CG13026, CG13031, CG13082, CG13239, CG13640, CG13731, CG13748, CG14096, CG14147, CG14572, CG14646, CG14661, CG14968, CG15098, CG15515, CG15545, CG15546, CG16884, CG17068, CG17224, CG17337, CG17549, CG17919, CG18067, CG2604, CG30101, CG30280, CG31626, CG31717, CG31812, CG32170, CG32579, CG32650, CG3301, CG33170, CG33946, CG34010, CG3552, CG3566, CG3700, CG42269, CG42568, CG42615, CG43367, CG4407, CG44286, CG45676, CG5390, CG5397, CG5867, CG5885, CG5934, CG6272, CG6429, CG6908, CG7457, CG7872, CG7896, CG8317, CG8319, CG8525, CG8788, CG9372, CG9521, CG9928, CHKov2, CIAPIN1, Cpr47Ea, Cpr49Aa, Cpr49Ah, Cpr64Aa, Cpr64Ad, Cpr78Cb, CR31144, CR43196, CR44107, CR44168, CR44264, CR44383, CR44468, CR44469, CR44621, CR44756, CR44888, CR44893, CR45150, CR45350, CR45501, CR45541, CR45682, CR45827, CR45828, CR45925, CR46108, CREG, Cyp6a15Psi, Cyp6d5, Cyp6g1, Cyp9f2, **Dip-B,** DnaJ-60, E(spl)mbeta-HLH, Ef1beta, Fmo-2, FucT6, Gapdh1, GstE8, GstO1, GstO2, GstO2, GstZ1, Hsf, Hsp70Bb, Lcp65Aa, Lsm10, mir-282, Mmp1, mRpS25, Myb, ND-B16.6, Nxt1, Obp56d, obst-A, olf186-M, Osi19, PCB, Pcd, PGRP-LF, pinta, Prosbeta3, Prx2540-1, Ptr, rempA, RpII15, RpS15Ab, scu, snmRNA:838, snoRNA:Psi18S-301, Syx8, tal-1A, tal-2A, tal-3A, tal-AA, Taldo, tRNA:CR32289, tRNA:H:48F, Ugt35b, vajk-1, ver, wal, yip2, yip2 **T3:** Acox57D-p, Acp54A1, Adk2, B-H1, Cdk7, CG10257, CG10344, CG10602, CG10754, CG11137, CG11191, CG11668, CG11852, CG12118, CG13082, CG13299, CG13731, CG13751, CG13928, CG14096, CG14572, CG14968, CG15515, CG15545, CG17337, CG17549, CG30280, CG32039, CG32170, CG3226, CG32579, CG32650, CG32806, CG32815, CG33170, CG33946, CG3566, CG3939, CG42239, CG44286, CG5397, CG5867, CG6272, CG6429, CG7457, CG7872, CG8500, CG8788, CG9319, CG9521, CG9928, CG9960, CR43700, CR44371, CR44420, CR45008, **CR45172**, CR45350, CR45554, CR45911, CR45919, CR45925, CR46108, CREG, Cyp6a15Psi, Ef1alpha100E, Faf2, Flo1, Fmo-2, Gapdh1, GstO2, Hsp70Bb, jdp, mei-S332, mir-282, mRpL27, mRpL35, mRpS25, ND-B16.6, Nup54, Obp56d, olf186-M, pinta, pncr015:3L, Prx2540-1, Pu, put, rempA, scu, Snapin, snoRNA:lola-a, snoRNA:Psi18S-301, Syx8, Taldo, Tes, TRAM, tRNA:CR32287, tRNA:CR32289, tRNA:H:48F, Vps4, yip2 ***bold similar in this study**
Pacifico et al, 2018 Time: 5-, 20-, 30-, and 40-day-old Tissue: Sex- and age-matched brains Condition: Age-related cognitive decline, aversive shock olfactory STM, 1min (sequential odors spaced 30sec) Technique: RNA-seq Analysis: STAR, HTSeq-count, DESeq2 with default filtering and WGCNA	**Male, likelihood ratio test (LRT) q-value < 0.05 (1076 genes identified):** too many to list here Similar in this study: **aay, aurA, ben, CG11425, CG13607, CG14879, CG1673, CG42394, CG5174, CG5446, CG5953, CG9098, cype, Gpdh, Rpl1, Sodh-2, Thor, TM4SF, TotA** ***bold similar in this study**
Signor and Nuzhdin, 2018 Time: 10, 20, 30min post exposure Tissue: Male and female heads, heterozygous inbred lines Condition: Ethanol vapor exposure from food Technique: RNA-seq Analysis: Exon delineation, a priori gene list filtering and abundance cutoffs	**Genes associated with exonic P_Treatment < 0.05 (1941 unique genes of 2402 identified):** too many to list here Similar in this study: **CG12116, CG13607, CG17816, CG2911, CG34166, CG3505, CG42540, CG9098, Crk, dor, endos, fus, jnj, l(2)k09913, mod(mdg4), NaCP60E, odd, Pitslre, Rpl1, Sec31, su(r), Trpm** ***bold similar in this study**
Other important studies with different methodologies: **Kong et al, 2010 (microarray); Morozova et al, 2009 (microarray); Morozova et al, 2015 (QTL analysis)**.	