Clemson University

TigerPrints

All Dissertations

Dissertations

8-2023

# A Systems Genetics Approach to Drosophila melanogaster Models of Rare and Common Neurodevelopmental Disorders

Rebecca MacPherson
rajones@g.clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Part of the Genetics Commons, and the Genomics Commons

## Recommended Citation

MacPherson, Rebecca, "A Systems Genetics Approach to Drosophila melanogaster Models of Rare and Common Neurodevelopmental Disorders" (2023). *All Dissertations*. 3386.
https://tigerprints.clemson.edu/all_dissertations/3386

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

# A SYSTEMS GENETICS APPROACH TO *DROSOPHILA MELANOGASTER* MODELS OF RARE AND COMMON NEURODEVELOPMENTAL DISORDERS

---

A Dissertation
Presented to
the Graduate School of
Clemson University

---

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Genetics

---

by
Rebecca Anne MacPherson
August 2023

---

Accepted by:
Dr. Trudy F. C. Mackay, Committee Chair
Dr. Robert R. H. Anholt, Co-Chair
Dr. Frank Alexander Feltus
Dr. Jennifer M. Mason

# ABSTRACT

Fetal Alcohol Spectrum Disorders are a group of disorders resulting from prenatal alcohol exposure, presenting with neurodevelopmental and facial abnormalities of varying severity. SSRIDDs and CdLS are rare disorders of chromatin modification, resulting in patients with a wide range of craniofacial, digit and/or neurodevelopmental abnormalities. All of these disorders have a wide range of clinical phenotypes and disease severity, yet the role of potential genetic modifiers and gene-gene or gene-environment interactions in disease pathogenesis is largely unknown and cannot be studied in humans. Insufficient numbers of patients with a single rare disorder prevent investigation of genetic factors beyond the focal disease-associated variant, while experimental study of the more common FASD using human subjects is prohibited due to ethical constraints. *Drosophila melanogaster* is an excellent model system for neurodevelopmental disorders, as Drosophila neurobiology is largely conserved in humans and experiments performed in Drosophila are low-cost, easily controlled, and exempt from regulation. Here, we take advantage of the Drosophila model system and identify genetic factors contributing to these neurodevelopmental disorders. Specifically, we used the Drosophila Genetic Reference Panel (DGRP) of inbred lines with full genome sequences and single cell RNA sequencing to identify genetic networks in adult Drosophila after developmental ethanol exposure and demonstrate that changes in sleep, activity, and time to sedation as a result of the developmental ethanol exposure are dependent on genetic background. We also developed a novel assay measuring time to ethanol-induced sedation of individual flies to better assess this phenotype in our research and characterized a previously unstudied long

noncoding RNA critical for Drosophila fitness and stress-response. We then established Drosophila models for multiple SSRIDD and CdLS subtypes and determined the extent to which behavioral and transcriptomic phenotypes vary within and across these rare disorders. Finally, we used SSRIDD Drosophila models to present evidence for the role of genetic modifiers in *ARID1B*-associated SSRIDD and identify candidate genetic modifiers for multiple SSRIDD subtypes. Taken together, these results show that the Drosophila model system is a powerful tool for investigating the genetic underpinnings of both rare and common neurodevelopmental disorders that cannot be currently identified using human populations.

# DEDICATION

I would like to dedicate this dissertation work to my family. To my husband, Grant, thank you for your encouragement, love, and understanding. You have unconditionally believed in me from the beginning, no matter the challenges we have faced, and have sacrificed much for both of our doctorates. To my son, John, in making me a mom, you have made me stronger, more patient, and more loving. You are my greatest accomplishment.

Thank you to my parents who enabled my scientific curiosity from a young age and for always truly being there for me. Thank you also to my sister who always cheers me on, even for my smallest achievements, and to my in-laws, for your never-ending enthusiasm. I could not have gotten to where I am without everyone’s support.

# ACKNOWLEDGMENTS

I would like to thank my advisors Dr. Trudy Mackay and Dr. Robert Anholt. You have fostered a collaborative and positive research environment at the CHG, and have supported my pursuits in education and outreach, as well as clinical genetics, providing me with professional development and mentoring along the way. Thank you too to my committee members Drs. Jennie Mason and Alex Feltus – I am so lucky as to have had such professional support throughout my Ph.D., contributing to a thoroughly positive experience. Also, thank you to those who wrote letters of recommendation on my behalf throughout my graduate career – your words directly contributed to my success.

Thank you to all my fellow lab members during my time in the Mackay-Anholt lab, whether past or present. A special thanks to my co-authors and collaborators, but also to anyone who helped me just for the sake of being kind; long days in the lab were made better with your company.

I would also like to thank the Greenwood Genetic Center for enriching my graduate experience beyond my own laboratory and for inspiring my research projects and career aspirations. A special thank you to Dr. Leta Tribble and the Division of Education for allowing me to volunteer my time and be a part of your summer programming.

# TABLE OF CONTENTS

Page

TITLE PAGE ..... i

ABSTRACT ..... ii

DEDICATION ..... iv

ACKNOWLEDGMENTS ..... v

LIST OF TABLES ..... ix

LIST OF FIGURES ..... xii

LIST OF SUPPLEMENTARY DATA FILES ..... xvi

CHAPTER

I. INTRODUCTION ..... 1

- Strengths of *Drosophila melanogaster* as a model system ..... 3
- *Drosophila melanogaster* as a model for alcohol-related phenotypes ..... 13
- *Drosophila melanogaster* as a model for rare neurodevelopmental disorders of chromatin modification ..... 27
- Research Purpose ..... 51
- References ..... 55

II. HIGH-THROUGHPUT METHOD FOR MEASURING ALCOHOL SEDATION TIME OF INDIVIDUAL *DROSOPHILA MELANOGASTER* ..... 81

- Author Contribution Statement ..... 81
- Introduction ..... 81
- Protocol ..... 84
- Representative Results ..... 91
- Discussion ..... 95
- Acknowledgements ..... 97
- References ..... 98

III. DEVELOPMENTAL ALCOHOL EXPOSURE IN DROSOPHILA: EFFECTS ON ADULT PHENOTYPES AND GENE EXPRESSION IN THE BRAIN....................................101

Author Contribution Statement............................................................101
Introduction..............................................................................101
Methods...................................................................................103
Results...................................................................................108
Discussion................................................................................119
Acknowledgements..........................................................................123
References................................................................................123

IV. MODULATION OF THE DROSOPHILA TRANSCRIPTOME BY DEVELOPMENTAL EXPOSURE TO ALCOHOL....................129

Author Contribution Statement............................................................129
Introduction..............................................................................129
Results...................................................................................132
Discussion................................................................................146
Methods...................................................................................149
Acknowledgements..........................................................................156
References................................................................................157

V. PLEIOTROPIC FITNESS EFFECTS OF THE LNCRNA *UHG4* IN *DROSOPHILA MELANOGASTER*.....................................162

Author Contribution Statement............................................................162
Introduction..............................................................................162
Results...................................................................................165
Discussion................................................................................182
Methods...................................................................................185
Acknowledgements..........................................................................196
References................................................................................197

VI. GENETIC AND GENOMIC ANALYSES OF *DROSOPHILA MELANOGASTER* MODELS OF CHROMATIN MODIFICATION DISORDERS........................................................208

Author Contribution Statement............................................................208
Introduction..............................................................................208
Results...................................................................................211

Discussion ........ 233
Methods ........ 236
Acknowledgements ........ 245
References ........ 246

VII. IDENTIFICATION OF CANDIDATE GENETIC MODIFIERS IN A *DROSOPHILA MELANOGASTER* MODEL OF *ARID1B*-RELATED SWI/SNF-RELATED INTELLECTUAL DISABILITY DISORDER ........ 254

Author Contribution Statement ........ 254
Introduction ........ 254
Methods ........ 258
Results ........ 265
Discussion ........ 273
Acknowledgements ........ 278
References ........ 278

VIII. DISCUSSION AND CONCLUSIONS ........ 283

References ........ 293

APPENDICES ........ 298

A: Supplementary material for Chapter 2 ........ 299
B: Supplementary material for Chapter 3 ........ 301
C: Supplementary material for Chapter 4 ........ 308
D: Supplementary material for Chapter 5 ........ 314
E: Supplementary material for Chapter 6 ........ 326
F: Supplementary material for Chapter 7 ........ 336

# LIST OF TABLES

Table Page

2.1 Measurements of ethanol sedation times (s) of individual flies of (A) DGRP lines RAL_177 and (B) RAL_555 for separate sexes (n = 48) .......... 92

2.2 Analyses of variance for sedation time across sex and DGRP line .......... 94

S2.1 Materials list .......... 300

S3.1 ANOVA tables for viability, ethanol sensitivity, sleep, and activity .......... 305

S3.2 Sequencing statistics .......... 305

S3.3 Genes used to annotate cell clusters .......... 305

S3.4 List of differentially expressed genes in each cluster in males .......... 305

S3.5 List of differentially expressed genes in each cluster in females .......... 306

S3.6 Human orthologs of differentially expressed genes .......... 306

S3.7 Common differentially expressed genes upon developmental alcohol exposure and acute exposure to cocaine .......... 307

S3.8 Comparison of cell type-specific differentially expressed genes between developmental ethanol exposure and acute cocaine exposure .......... 307

S3.9 Gene ontology analysis of differentially expressed genes identified both after developmental exposure to alcohol and acute intake of cocaine .......... 307

S5.1 Analyses of behavioral phenotypes .......... 323

List of Tables (Continued)

Table Page
S5.2 Differentially expressed genes ........ 323

S5.3 Analysis of differential expression ........ 323

S5.4 Enriched Gene Ontology (GO) Biological Process, Molecular Function and REACTOME terms and their raw and BH-FDR adjusted *p*-values ........ 323

S5.5 Lists of gene names, symbols, and Flybase IDs (FBID) for each network cluster and k-means cluster ........ 324

6.1 Drosophila genes used in fly models ........ 213

6.2 Differentially expressed gene counts ........ 224

S6.1 Fly reagents and primer sequences ........ 331

S6.2 Ortholog prediction scores for potential focal genes ........ 331

S6.3 Percent knockdown of focal genes ........ 331

S6.4 Quantification of changes in behavior and brain morphology from knockdown of focal genes ........ 331

S6.5 ANOVA results from differential expression analyses ........ 332

S6.6 Overlap of differentially expressed genes across analyses ........ 332

S6.7 k-means threshold ........ 332

S6.8 k-means clustering gene lists ........ 333

S6.9 Gene Ontology (GO) analyses for differentially expressed genes ........ 333

S6.10 Gene Ontology (GO) analyses for k-means clusters ........ 333

S6.11 Ortholog prediction scores for differentially expressed genes ........ 333

S6.12 Ortholog prediction scores and known disease associations for co-regulated genes ........ 333

List of Tables (Continued)

Table Page
S6.13 Percent knockdown of co-regulated genes ............................................... 334

S6.14 Quantification of changes in behavior from knockdown of co-regulated genes ............................................................................ 334

S7.1 Drosophila stock reagents ......................................................................... 340

S7.2 Quantification of *osa* knockdown ............................................................ 340

S7.3 Analysis of sleep and activity phenotypes from knockdown of *osa*. ....................................................................................... 340

S7.4 Analysis of background-dependent changes in sleep and activity phenotypes from knockdown of *osa*. ..................................... 340

S7.5 Line means for wing vein scores from knockdown flies ........................... 341

S7.6 Vein counts for control flies ..................................................................... 341

S7.7 Analyses of variance and broad sense heritability for *MS1096>osa* x DGRP flies ................................................................. 341

# LIST OF FIGURES

Figure Page

1.1 Select components of the *D. melanogaster* brain ........................................ 7

2.1 Diagram of the testing apparatus and filming chamber .............................. 85

2.2 Photograph of the assay system ................................................................. 86

2.3 A fly aspirator in which flies are collected with an interchangeable mouthpiece attached to flexible tubing and a wide bore serological pipette with a cotton gauze stopper .................................... 88

2.4 Alcohol sedation times of DGRP lines RAL_177 and RAL_555 ............... 93

3.1 Diagram of the experimental design ......................................................... 104

3.2 Effects of developmental alcohol exposure on viability and behavioral phenotypes in adult flies .................................................... 109

3.3 Uniformity across samples of single cell transcriptomes .......................... 112

3.4 UMAP visualization and annotation of cell clusters.................................. 113

3.5 Differentially expressed genes across clusters in males (A) and females (B) after developmental alcohol exposure ....................... 115

3.6 Global interaction networks of differentially expressed gene products in males (A) and females (B) following developmental alcohol exposure ......................................................... 117

3.7 Venn diagrams indicating the proportions of differentially regulated genes after exposure to alcohol during development or acute consumption of cocaine for males (A) and females (B) ........ 119

S3.1 Actograms showing average number of counts per fly per minute ......................................................................................... 302

S3.2 Differentially expressed genes across clusters in males (A) and females (B) after developmental alcohol exposure ....................... 303

List of Figures (Continued)

Figure Page

4.1 Correlations of differences in gene expression between developmental ethanol treatment and control in females ....................135

4.2 Correlations of differences in gene expression between developmental ethanol treatment and control in males........................137

4.3 Female (A) and male (B) interaction networks built from eQTLs and known genetic and physical associations..........................140

4.4 Differentially expressed snoRNAs from female flies grown on ethanol versus regular food............................................................143

4.5 Effects of developmental ethanol exposure on sleep and activity phenotypes ..............................................................................145

S4.1 Correlations between variation in expression of snoRNAs after chronic exposure to ethanol and variation in expression of their host genes ..................................................................309

S4.2 Correlations between ethanol-induced variation in expression of snoRNA host genes and variation in expression of *CycE* ...............310

5.1 *Uhg4* deletions ..........................................................................166

5.2 Late-stage oocytes are absent in *Uhg4* deletion lines ...............................168

5.3 Effects of *Uhg4* deletion on egg-adult development and viability ...........170

5.4 Effects of *Uhg4* deletion on responses to stress........................................173

5.5 Effects of *Uhg4* deletion on sleep and activity phenotypes.......................176

5.6 Interaction networks of genes coregulated with *Uhg4* .............................181

S5.1 Crossing scheme to generate *Uhg4* deletion CRISPR mutants .................315

S5.2 Mating phenotypes of *Uhg4* deletion flies.................................................317

S5.3 Non-monotonic relationship between raw and Benjamini-Hochberg adjusted p–values ..............................................................319

List of Figures (Continued)

Figure Page

S5.4 K-means clusters derived from genes and NTRs with differential expression in a global transcriptomic analysis using genes significant for the *Line* or *Line*×*Sex* terms (BH-FDR < 0.1). .............. 321

S5.5 Additional interaction networks ................................................................ 322

6.1 Altered startle response phenotypes in SSRIDD and CdLS fly models ........................................................................................ 214

6.2 Altered sleep and activity phenotypes in SSRIDD and CdLS fly models ........................................................................................ 216

6.3 Examples of mushroom body abnormalities in SSRIDD and CdLS fly models ............................................................................ 218

6.4 SSRIDD and CdLS fly models show gene-specific changes in mushroom body and ellipsoid body ................................................ 220

6.5 k-means clusters for females ....................................................................... 226

6.6 k-means clusters for males .......................................................................... 228

S6.1 Gross viability observations in potential CSS/NCBRS and CdLS fly models ......................................................................................... 327

S6.2 Selection of genes for k-means clustering ................................................ 328

S6.3 Overlap of differentially expressed genes in SSRIDD and CdLS fly models ......................................................................................... 329

S6.4 Altered phenotypes due to knockdown of co-regulated genes .................. 330

7.1 Altered sleep and activity phenotypes in a Drosophila *ARID1B*-SSRIDD model ................................................................... 266

7.2 Background-dependent sleep and activity changes in Drosophila *ARID1B*-SSRIDD models ................................................................... 268

7.3 Examples of wings with altered veins ....................................................... 270

List of Figures (Continued)

Figure Page
7.4 Line means for wing vein scores ............................................................ 273

S7.1 Crossing scheme to generate *Ubi156>osa* flies........................................ 337

S7.2 Crossing scheme to generate *MS1096>osa* flies ....................................... 338

S7.3 Correlation of line means across sexes ...................................................... 339

# LIST OF SUPPLEMENTARY DATA FILES

File Page

S4.1 DGRP lines used for transcriptional profiling ..........................................312

S4.2 Three way mixed model ANOVA for differences in transcript abundances between flies reared on regular and ethanol supplemented medium for sexes combined ........................................312

S4.3 ANOVA for differences in transcript abundances between flies reared on regular and ethanol supplemented medium for (A) females and (B) males ..........................................................312

S4.4 Genes with significant *LxT* interactions in transcript abundances.............312

S4.5 (A) eQTL mapping and model selection results for genes with a statistically significant *L*x*T* ANOVA term in females and males. (B) Genes containing eQTLs (in the gene body or within a 1,000 bp window of the eQTL) in females and males............................................................................313

S4.6 snoRNAs with altered gene expression following chronic exposure to ethanol during development and their host genes ............313

S4.7 Mixed model ANOVAs for sleep and activity phenotypes .......................313

S5.1 Video of a representative *Uhg4* deletion (*208-ΔG*) female fly displaying an elevated, splayed resting wing position compared to a wildtype female fly......................................................325

S5.2 Video of a representative *Uhg4* deletion (*208-ΔG*) male fly displaying an elevated, splayed resting wing position compared to a wildtype male fly.........................................................325

S6.1 Video of tapping behavior in a male fly with knockdown of *vtd* following a startle response ......................................................335

S6.2 Video of control male fly following a startle response..............................335

# CHAPTER ONE

# INTRODUCTION

Biology is driven by genetics, the environment, and the interaction between the two. However, insurmountable obstacles prevent scientists from fully understanding processes influenced by human genetics. Ethical constraints prohibit anything but the best treatment of individuals, monetary limits prevent study of sufficient quantities of people to make confident assertions, and long lifespans and generation times for humans prevent collection of longitudinal data and generation of large numbers of offspring for study. However, as biology is conserved across taxa, model organisms can fill knowledge gaps and provide starting points for future studies using human data. From yeast to apes, each model system has its own strengths and pitfalls. Model organisms allow for detailed investigation of biological concepts relevant to human health, with fewer regulations, at lower cost, and in a shorter timeframe than studies on humans alone. Furthermore, genetic and environmental factors can be controlled and even manipulated in experiments using model systems, whereas in humans, these factors cannot often be altered.

Genetic model systems also enable a cost-effective systems genetics approach to complex traits. In this era of multi-omics, one project can generate transcriptome, metabolome, proteome, epigenome, microbiome and/or behavioral data across a variety of genetic backgrounds and environments. Taking a systems genetics approach, with integration of all available data, gives a more complete picture of the molecular underpinnings of a trait of interest (Civelek and Lusis 2014). A systems genetics approach is not limited to traditionally complex traits; even seemingly simple monogenic and/or Mendelian disorders are more complicated than previously realized and could be considered a complex trait (Scriver and Waters 1999; Bis-Brewer *et al.* 2020). This holistic systems genetics approach can generate genetic networks for and identify molecular correlates with the trait of interest, which could lead to identification of genetic modifiers, environmental interventions, and/or novel targets for pharmaceuticals.

In this dissertation, I utilize the model organism *Drosophila melanogaster* as a model for multiple human diseases, including Fetal Alcohol Spectrum Disorder (FASD), Switch-Sucrose non-fermenting (SWI/SNF)-Related Intellectual Disability Disorders (SSRIDDs), and Cornelia de Lange syndrome (CdLS). Using the models I have established, I take a systems genetics approach to investigate the role of genetic background in modulating disease severity. I begin with an introduction chapter on the strengths of Drosophila as a model system, then shift my focus to alcohol and how Drosophila has been used to understand the influence of genetics on consequences of developmental ethanol exposure. I conclude the introduction with an overview of rare neurodevelopmental disorders of chromatin modification and how Drosophila can be used to model these disorders and identify potential modulators of disease severity. In subsequent chapters of this dissertation, I describe a novel method for assessing ethanol sedation time of single Drosophila, identify genetic networks associated with genetic background-dependent responses to developmental ethanol exposure, characterize a long noncoding RNA critical for reproductive fitness and ethanol-mediated stress response, utilize single cell RNA sequencing to understand the effect of developmental ethanol exposure in the adult brain, establish Drosophila models for related neurodevelopmental disorders of chromatin modification and compare transcriptomic changes across Drosophila models, and analyze the role of non-additive epistasis in, and identify candidate genetic modifiers for Drosophila models of SWI/SNF-Related Intellectual Disability Disorders.

## Strengths of *Drosophila melanogaster* as a model system

*D. melanogaster* has been utilized as a genetic model organism for over 100 years (Morgan 1910) and is one of the premier model organisms in biology. *D. melanogaster* is a relatively simple eukaryote with *X* and *Y* sex chromosomes and two large (chromosomes *2* and *3*) and one very small (chromosome *4*) autosome , containing about 14,000 protein-coding genes. *D. melanogaster* was the first major higher-order organism to have its genome fully sequenced (Adams *et al.* 2000; Hoskins *et al.* 2015; Gramates *et al.* 2022). Furthermore, Drosophila have a short generation time, can be inbred or crossed, and can be rapidly phenotyped using an ever-growing number of methods. Drosophila are well suited to serve as a model organism for human disorders, as Drosophila have orthologs of about 65% of human disease genes (Reiter *et al.* 2001; Chien *et al.* 2002), are economical to rear in large numbers in a controlled environment, and are exempt from regulatory oversight. As a result, Drosophila have been used to model a range of human diseases, including substance use disorders, rare genetic diseases, neurological disorders, obesity, cancer, and sleep disorders (Ugur *et al.* 2016; Dubowy and Sehgal 2017; Mirzoyan *et al.* 2019; Petruccelli and Kaun 2019). Studies on human diseases are facilitated by publicly available *D. melanogaster* genetic tools, including RNA interference strains with knockdown, knockout, and/or overexpression constructs for most Drosophila genes that can be used in a tissue-specific or time-specific manner (Dietzl *et al.* 2007; Zirin *et al.* 2020), balancer chromosomes to prevent recombination and/or preserve deleterious mutations, and thousands of Drosophila strains with characterized alleles and mutations (Gramates *et al.* 2022). The availability of the *Drosophila melanogaster* Genetic Reference Panel (DGRP) of wild-derived inbred Drosophila lines with full genome sequences (Mackay *et al.* 2012; Huang *et al.* 2014) enables genome wide association (GWA) mapping of natural allelic variants affecting quantitative traits with high precision, since linkage disequilibrium decays rapidly with physical distance in regions of normal recombination.

### *The Drosophila brain and sleep*

The *D. melanogaster* nervous system is well-suited as a model for human disorders with a neurological component. Like humans, Drosophila have glia and neurons with similar functional roles as their human counterparts, utilize many of the same neurotransmitters (*e.g.* Gamma Aminobutyric Acid (GABA), dopamine, serotonin), and are capable of experience-dependent modulation of behavior (Quinn *et al.* 1974; Bellen *et al.* 2010; Li *et al.* 2020). While the complex eye and its development were one of the first systems used in Drosophila neurogenetics research, scientific progress has advanced to single-cell transcriptomic atlases of *D. melanogaster* brain tissue, with multiple time points and multiple genetic backgrounds (Davie *et al.* 2018; Allen *et al.* 2020; Janssens *et al.* 2022), as well as in-depth analyses of neural circuitry of Drosophila brain components, including the mushroom body (Li et *et al.* 2020). The mushroom body is comprised of about 2,000 neurons called Kenyon cells, and is a key regulator of sleep, experience-dependent modulation of behavior, associative memory, and integration of signaling from dopaminergic neurons (Joiner *et al.* 2006; Li *et al.* 2020; Modi *et al.* 2020). Downstream targets of the Kenyon cells are mushroom body output neurons (MBONs), and the synapses between Kenyon cells, dopaminergic neurons, and MBONs form the lobes of the mushroom body, *e.g* alpha and beta lobes. The morphological characteristics of the alpha and beta lobes vary with genetic background (Zwarts *et al.* 2015), yet abnormalities in lobe structure have also been observed in Drosophila models of human disorders with neurological manifestations (Chubak *et al.* 2019; Kang *et al.* 2019; Kim *et al.* 2021). The *D. melanogaster* brain has also been used as a model for neurodegenerative disorders, including Alzheimer’s disease, Parkinson’s disease, and Huntington’s disease (Bolus *et al.* 2020).

Drosophila is also an important model organism for sleep. Although Drosophila do not have states of rapid eye movement (REM) and non-REM sleep characteristic of higher-order organisms such as mammals (Ly *et al.* 2018), Drosophila sleep does satisfy the behavioral characteristics of sleep, including regulation by circadian rhythm, decreased responsiveness to stimuli, and rebound following deprivation (Campbell and Tobler 1984; Hendricks *et al.* 2000; Shaw *et al.* 2000). Furthermore, *D. melanogaster* have a distinct body posture for sleeping and sleep at a preferred location, with a threshold of about 5 minutes of immobility indicative of sleep (Hendricks *et al.* 2000). Sleep can be easily quantified using an activity monitor, which quantifies the number of times an individual fly breaks an infrared beam, thereby indicating activity, and the absence of activity for at least 5 minutes indicating sleep. Sleep is sexually dimorphic and is influenced by age; males sleep more than mated females during the day (Huber *et al.* 2004; Isaac *et al.* 2010), and aged Drosophila show more fragmented sleep patterns (Koh *et al.* 2006). Regulation of sleep has been associated with numerous genes and neurotransmitters, many of which are involved in regulation of dopamine, GABA, serotonin, glutamate, and calcium in the Drosophila brain. The Drosophila mushroom body is instrumental for regulation of sleep; without a functioning mushroom body sleep is altered (Joiner *et al.* 2006; Pitman *et al.* 2006). Specific MBONs and Kenyon cells have been classified as sleep-promoting or wake-promoting, further demonstrating the importance of the mushroom body in regulation of sleep (Ly *et al.* 2018). In addition to its role as a regulator of locomotor activity, the central complex is also responsible for regulation of sleep. In the fan-shaped body, dopamine released by dopaminergic neurons counteracts sleep-promoting neurons, leading to wakefulness (Pimentel *et al.* 2016). The sleep-promoting R2 ring neurons of the ellipsoid body become more excitable in response to sleep deprivation, but less excitable upon sleep rebound (Liu *et al.* 2016).

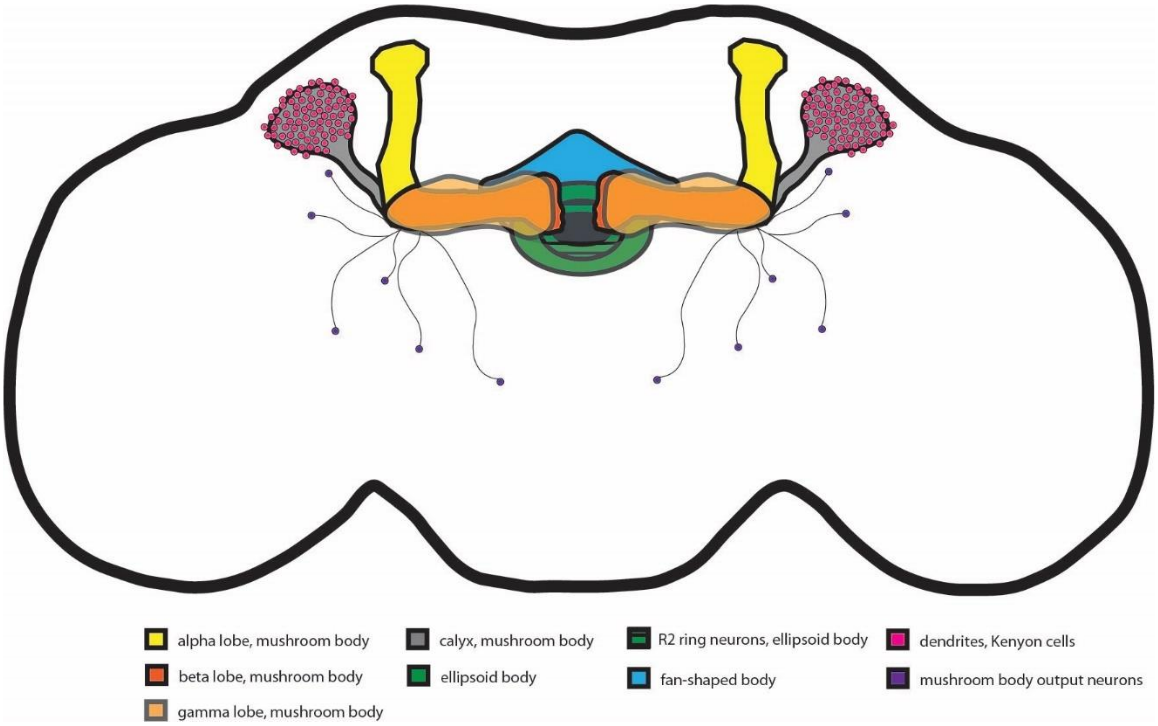

alpha lobe, mushroom body
beta lobe, mushroom body
gamma lobe, mushroom body
calyx, mushroom body
ellipsoid body
R2 ring neurons, ellipsoid body
fan-shaped body
dendrites, Kenyon cells
mushroom body output neurons

**Figure 1.1. Select components of the *D. melanogaster* brain**. Schematic includes mushroom body and central complex components important in regulation of sleep in Drosophila. Drawing not to scale.

*The Drosophila melanogaster Genetic Reference Panel*

While *D. melanogaster* has proven fruitful in studies on the effects of single genes and mutations with large effects. The use of Drosophila, with its strengths as a model organism, facilitates studies with adequate statistical power to begin to unravel the genetics underpinning variation in quantitative (or complex) traits. The DGRP has been used to perform genome-wide association studies, mapping key loci associated with variation for Drosophila quantitative traits, including heavy metal toxicity, fitness traits, mating behaviors, stress-response, and body morphology (Mackay and Huang 2018). The genetic variation observed for these traits supports hypotheses about the importance of genetic background and the potential influence of genetic modifiers on a variety of phenotypes. Recently, the DGRP has been expanded from about 200 lines to over 1000 lines (T. Mackay, personal communication, March 2023), all of which have been sequenced to an average depth of 40X using short-read next generation sequencing technology. A subset of 100 new DGRP lines have also been sequenced with long-read technology to accurately capture structural variants, mid-sized deletions, translocations, transposons, and highly repetitive areas of the genome ((T. Mackay, personal communication, March 2023), The larger number of DGRP lines with greater sequencing accuracy will provide greater power to detect epistatic interactions and variants with smaller effect sizes and or lower allele frequencies, as well as low frequency variants.

The DGRP has also been used to identify gene by environment interactions across the Drosophila genome (Mackay and Huang 2018; Huang *et al.* 2020). Introduction of foreign substances, (*e.g.* pharmaceuticals, heavy metals, ethanol) into Drosophila food media allows examination of effects of a specific substance on phenotypes via developmental exposure and/or consumption. Individuals within each DGRP line are nearly genetically identical, but there is wide genetic variation between DGRP lines, which gives excellent power to detect gene by environment interactions when large numbers of individuals from each line are reared in multiple controlled environmental treatments. Genetic modifiers for a specific genetic perturbation can also be identified using the DGRP (He *et al.* 2016, Talsness, *et al.* 2020; Ozsoy *et al.* 2021). Progeny resulting from crosses of a Drosophila line with a mutation or gene knockdown to numerous DGRP lines can be assessed for phenotypic variation, and subsequent GWA analyses can identify variants associated with the severity of the phenotype, e.g. variants that may be a genetic modifier of the original genetic perturbation.

### *CRISPR/Cas9 System in Drosophila*

One advantage of working with *D. melanogaster* as a model organism is the feasibility and ease of genome manipulation. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 system is a powerful and efficient tool in Drosophila to infer gene function, perturb regulatory networks, and generate precise single base pair changes or insertions or deletions of entire gene sequences. Generally, the endonuclease (Cas9 is commonly used) generates double stranded breaks 3 nucleotides away from a PAM site. The specific PAM site(s) used by the endonuclease is determined by the short guide RNA(s) (gRNA) designed by the researcher. Insertions can be generated by ensuring desired DNA fragment is present during repair of the double stranded break generated by the endonuclease. Although the potential for genome manipulation was first discovered in bacteria (Jinek *et al.* 2012), the CRISPR/Cas9 system was quickly adapted for *D. melanogaster* (Gratz *et al.* 2013). Since 2013, CRISPR has been widely utilized in *D. melanogaster* and nucleases with alternative properties (e.g. greater PAM site flexibility, increased specificity) continue to be implemented (Zirin *et al.* 2022). The endonuclease, gRNAs, and any DNA fragments can be supplied exogenously and co-injected into Drosophila embryos to generate mutations, but individual components (e.g. gRNAs) can be placed into the Drosophila genome and crossed with another transgenic line (e.g. one expressing *Cas9*) to yield progeny with potential mutations of interest (Zirin *et al.* 2022). Appropriate balancer chromosomes in *D. melanogaster* allow preservation of otherwise homozygous lethal or deleterious mutations for further study and prevent recombination from interfering with the specific mutation of interest. Although CRISPR can efficiently generate mutations at precise locations, the stochastic nature of DNA repair is evident across mutations isolated from the same experiment; there is still variation in sequence between isolated mutations and the possibility of off-target double-stranded breaks remains (Martin *et al.* 2016).

### *UAS-GAL4 system in Drosophila*

Another powerful component of the *D. melanogaster* genetic toolkit is the *UAS-GAL4* system. *GAL4* is a yeast transcriptional activator, which cooperatively binds to an Upstream Activating Sequence (*UAS*). The *UAS-GAL4* system has two primary constructs: a tissue-specific or ubiquitous enhancer, which is located upstream of the yeast transcription factor *GAL4,* and a *GAL4* binding site (*UAS* sequence) located upstream of a gene-specific sequence (Brand and Perrimon 1993). This gene-specific sequence, when transcribed, can be complementary in sequence to the gene of interest that leads to decreased gene expression due to RNA interference (RNAi), or short guide RNAs that induce increased gene expression when in a background also containing Cas9 protein (Heigwer *et al.* 2018). Additional variations on this system include temperature- and time-specific *GAL4* expression, often through the *GAL4* repressor protein *GAL80* to limit expression of *GAL4*, and using *GAL4* to drive expression of *Cas9* in the presence of a guide RNA to induce a CRISPR-mediated genomic alteration (Heigwer *et al.* 2018). Some of the primary pitfalls of the *UAS-GAL4* system in Drosophila are the potential for off target effects, inconsistent levels of gene knockdown, and negative impacts on fitness (Ma *et al.* 2006; Alic *et al.* 2012; Vissers *et al.* 2016).

In addition to a large number of publicly available *GAL4* drivers (Jenett *et al.* 2012; Gramates *et al.* 2022), there are multiple collections of lines with *UAS* constructs located on chromosomes *X*, *2*, and/or *3* for almost all *D. melanogaster* genes. While all of these collections have the potential for off-target effects, the Vienna Drosophila Research Center "GD" collection is limited in value because the *UAS*-construct insertion sites for these lines are unknown (Dietzl *et al.* 2007; Green *et al.* 2014; Vissers *et al.* 2016; Evangelou *et al.* 2019). The RNAi lines available as part of the Transgenic RNAi Project (TRiP), generated by the Drosophila RNAi Screening Center (DRSC), are isogenic and have a single known *UAS* insertion site for each chromosome (*attp40* and *attp2* for chromosomes *2* and *3*, respectively) (Perkins *et al.* 2015; Zirin *et al.* 2020). For some genes, the TRiP collection has two RNAi lines, one with the *UAS* construct on chromosome 2 and the other with the construct on chromosome *3*. The ability to select between two isogenic RNAi lines for the same gene of interest provides the researcher the opportunity to use both RNAi lines independently for stronger evidence of a finding and the ability to select the *UAS* construct on the chromosome that works best in crossing schemes used to generate specific genotypes.

Development of new genetic resources, such at those outlined above, ensure Drosophila remains one the of the primary model systems in biology. However, many of the novel technologies, including variations on single-cell technologies, tissue preservation, and genome manipulation are developed with human and/or mouse samples in mind. Implementation of these methods on alternative tissue types, such as those found in Drosophila, is challenging and typically requires protocol development and validation by the individual labs that wish to use the technology.

Although the existence of stock centers allows widespread distribution of Drosophila stocks, decades-long maintenance of stocks, despite a lack of contamination, are still susceptible to mutations and genetic drift. The degree of drift and divergence in individual Drosophila stocks is rarely quantified, thus some stocks are not what they seem. Furthermore, mutations and RNAi collections are initially created in different genetic backgrounds, which can result in two studies on the same gene or variant having seemingly contradictory outcomes. Comparisons of multiple members of a single protein complex stemming from research that was not performed in isogenic backgrounds may not be as accurate as initially proposed. Much of the field does not appreciate the potential magnitude of background-specific effects, especially in single-gene studies, though continued development and use of stock libraries containing isogenic lines for genome- and transcriptome-manipulation may help alleviate some of the genetic-background-based inconsistencies going forward.

## *Drosophila melanogaster* as a model for alcohol-related phenotypes

### *Alcohol Use in Humans*

Ethanol is the most widely consumed and abused drug. In the U.S., an estimated 5.6% of adults have Alcohol Use Disorder (AUD), which is a decreased ability to control alcohol use in spite of the deleterious consequences of such use (Substance Abuse and Mental Health Services Administration 2020; National Institute on Alcohol Abuse and Alcoholism 2021). Individuals with AUD may also develop tolerance to ethanol and require more ethanol to feel similar effects. This tolerance can be metabolic, where liver enzymes metabolize alcohol more rapidly; functional, where the brain can compensate for the relatively high levels of ethanol in the body; or even learned, where an individual's tolerance is dependent on environmental cues (Elvig *et al.* 2021). Several factors influence the risk of developing the brain disorder AUD, including mental health comorbidities, early drinking age, ethanol exposure during brain development, alcohol misuse, and family history (Hasin and Grant 2015). Estimates of AUD heritability are as high at 60% (Tawa *et al.* 2016), indicating the strong role of genetics, though environment and the interactions between an individual's genes and their environment also influences the likelihood of developing AUD.

In humans, ethanol crosses the blood-brain barrier and acts as a central nervous system depressant, enhancing GABA-A receptor mediated inhibition and antagonizing excitatory N-methyl-D-aspartate (NMDA) glutamate receptors. Imaging studies have shown that ethanol can affect numerous brain regions, including the nucleus accumbens, dorsolateral and mesial prefrontal cortex, amygdala, and the dorsolateral, inferior prefrontal and parietal cortices (Bjork and Gilman 2014). Negative consequences of ethanol use are not limited to AUD, as excessive alcohol consumption, even in the absence of AUD, is associated with liver cirrhosis, cancer, violence, and accidents and injuries (National Center for Chronic Disease Prevention and Health Promotion 2022). Females that are pregnant or may become pregnant who consume alcohol put a developing fetus at risk of Fetal Alcohol Spectrum Disorder (FASD).

### *Fetal Alcohol Spectrum Disorders in Humans*

FASDs refer to a range of conditions stemming from prenatal alcohol exposure. Although FASD affects 0.77% of the population worldwide (Lange *et al.* 2017), the rate of FASD is higher in the U.S., where up to five percent of the school aged population may have FASD (May *et al.* 2014). The most severe form of FASD is Fetal Alcohol Syndrome (FAS), which is characterized by intellectual disability and learning problems, decreased head size, height, and weight, as well as distinct facial features including a smooth philtrum and short palpebral fissures (May *et al.* 2014). Individuals with FAS present a financial burden on their caretakers with a cost of about 2 million per individual, leading to an estimated cost of $4 billion annually in the U.S. alone (Lupton *et al.* 2004).

FASDs are completely preventable if the fetus is not exposed to alcohol during gestation. However, a 2022 study found that 1 in 7 pregnant women drank at least one alcoholic beverage in the 30 days prior to survey administration, and 1 in 20 pregnant women reported binge drinking (Gosdin *et al.* 2022). The effects of prenatal alcohol exposure are difficult to quantify since women who do not know they are pregnant may consume alcohol, potential comorbidities, issues associated with self-reporting of alcohol use during pregnancy, and inability to quantify the environment and degree of alcohol exposure for the fetus. Furthermore, alcohol metabolism and alcohol use disorder have heritable components (Dodge *et al.* 2014; Tawa *et al.* 2016) and are therefore dependent upon genetic background, further skewing conclusions made from observational studies on prenatal ethanol exposure.

A twin study comparing diagnosis rates of FASD found that diagnosis was 100% concordant in monozygotic twins, but only about 65% concordant in dizygotic twins, suggesting a genetic component to susceptibility to the effects of prenatal alcohol exposure (Streissguth and Dehaene 1993). Ethanol metabolism could contribute to background-dependent ethanol susceptibility, as specific alcohol dehydrogenase alleles are significantly underrepresented in children unaffected with FAS, correlated with lower FASD incidence (McCarver *et al.* 1997; Viljoen *et al.* 2001; Das *et al.* 2004; Jacobson *et al.* 2006), and associated with specific ethanol-associated phenotypes (Boyles *et al*. 2010), including alcohol intake of an adult (e.g., a pregnant mother). GWA studies and candidate gene approaches focusing on cleft lip/palate have identified SNPs associated with ethanol use and facial clefting (Romitti *et al.* 1999; Beaty *et al.* 2011), supporting the hypothesis that genetics influences phenotypic presentation of FASD as well.

Overall, little is understood about the exact genetic underpinnings and pathogenic mechanisms of developmental alcohol exposure. Detailed analyses on the molecular and genetic effects of ethanol on brain tissue are extremely limited, given the difficulty in obtaining brain tissue from living individuals. Furthermore, information must be obtained in a retrospective manner, as it is unethical to administer a neurotoxic substance. Thus, our knowledge of the transcriptomic and epigenetic effects of ethanol on human brain is derived from post-mortem tissues. Although post-mortem tissue may be the best available human brain tissue, it does not provide any molecular information on acute effects, effects during neonatal and childhood development, and effects due to repeated exposures to alcohol. As previously discussed, human data can only provide observational findings – experiments using human data, outside of human cell lines (which also have limited translatability to humans on a macroscale), for studies on humans and ethanol are not possible, regardless of the question of interest. Model organisms provide a unique opportunity to investigate the effects of ethanol, especially during development.

### *Mammalian models of Fetal Alcohol Spectrum Disorder*

One theme across multiple animal models of FASD is the influence of strain-specificity on experimental outcomes. Despite a controlled laboratory environment, different genotypes can result in variable phenotypes, whether in mouse, rat, chicken, Drosophila, or zebrafish models of FASD (Eberhart and Parnell 2016). In response to developmental ethanol exposure, genetic background-specific differences have been observed in craniofacial dysmorphology, survival, apoptosis of neural crest cells, eye morphogenesis, overall birth defects, and cerebellar changes (Eberhart and Parnell 2016). Interestingly, even sub-strains of rat responded differently to developmental ethanol exposure (Wentzel and Eriksson 2008), further highlighting the impact that small genetic differences may have on phenotype. The strain-specificity complicates examination of the molecular mechanisms of FASD and emphasizes the need for controlled genetic backgrounds in future studies.

Animal models are also invaluable for assessing long-terms impacts of developmental alcohol exposure. Much of the research on FASD is focused on direct effects of ethanol on development, with a few studies on children and adolescents and almost no information about outcomes for FASD patients as adults. Thus, the long-term health impacts of developmental ethanol exposure remain understudied. Although mental health disorders stemming from developmental ethanol exposure are challenging to model outside of humans, animal models of FASD have found that developmental ethanol exposure may predispose individuals to hypertension, type II diabetes, cancer and tumorigenesis, substance abuse, and autoimmune disorders (Moore and Riley 2015). Animal models have also found that alcohol exposure in adults can impact ethanol sensitivity across at least two subsequent generations (Bonilla *et al.* 2021; Guzman *et al.* 2022), studies which are nearly impossible to conduct in human populations.

Much of the work on craniofacial effects of developmental ethanol exposure has stemmed from work in zebrafish, especially given the ease of controlled alcohol exposure during development. For example, work done in zebrafish indicates that various components of the *Sonic Hedgehog* pathway, which is known to cause holoprosencephaly in humans, interact with ethanol to cause craniofacial and neurological changes during development (McCarthy *et al.* 2013; Zhang *et al.* 2013), some of which can be ameliorated by cholesterol (Li *et al.* 2007). Specific mutant mouse strains showed exacerbation of holoprosencephalic phenotypes in response to developmental alcohol exposure that were rescued by mutations in genes that increased Shh signaling (Hong and Krauss 2012, 2013). Alongside other non-mammalian model systems, including chicken, quail and frog, fish models of FASD show developmental ethanol exposure interferes with signaling of retinoic acid, a key morphogen which also relies upon ADH for proper catabolism (Moore and Riley 2015).

Mammalian models of prenatal alcohol exposure are well suited for modeling effects on neuronal tissues, the presence of specific gestational timeframes, and the ability of the mammalian placenta and uterus to best mimic the human fetal environment. Although mouse models have led to mechanistic findings about the role of ethanol during development, such as the importance of *NOS1* in brain development in the presence of developmental ethanol exposure (Bonthius *et al.* 2015; Karacay *et al.* 2015A; Karacay *et al.* 2015B), mouse models generally show decreased brain volume as a result of early gestational ethanol exposure, findings which are also true of human FASD patients (Petrelli *et al.* 2018). Mouse models of developmental ethanol exposure in later gestation have also shown changes in organ development, somatosensory, visual and frontal cortexes, delayed developmental milestones, and increased depressive and anxiety behaviors, all of which are similar to phenotypes observed in FASD patients (Petrelli *et al.* 2018). Mouse FASD models have also shown decreased short-wave sleep and increased sleep fragmentation (Wilson *et al.* 2016), which align with sleep disturbances reported in FASD patients (Inkelis and Thomas 2018).

Experiments using animal models of FASD and AUD can have controlled environments, ethanol dosage, and genetic backgrounds. Furthermore, ethanol metabolism and genes related to the nervous system are highly conserved and mammalian models can emulate the *in utero* ethanol exposure of humans. Although mammalian models have been primarily used in this field, these models suffer from small sample sizes, limited genetic resources, and some governmental regulations. Thus, large scale “-omic” work, effects of genetic background, and comprehensive whole-brain research regarding ethanol exposure are largely understudied and under-sampled and remain relative knowledge gaps in the field.

### *Drosophila as a model system for alcohol-related phenotypes*

*D. melanogaster* has been used in numerous studies of the effects of drugs, including ethanol. Drosophila and humans both metabolize ethanol though alcohol dehydrogenase and acetaldehyde dehydrogenase enzymes, and developmental ethanol exposure can lead to changes in development and adult behavior (McClure *et al.* 2011; Devineni and Heberlein 2013; Morozova *et al.* 2018). As in humans, the Drosophila CNS is stimulated at low doses of ethanol but depressed at higher doses of ethanol (Devineni and Heberlein 2013). These effects on the nervous system manifest in a similar manner to humans: low doses of ethanol stimulate locomotor activity and higher doses lead to sedation and even death. Drosophila also develop tolerance and preference with repeated exposures to ethanol and experience withdrawal when the source of ethanol is removed (Berger *et al.* 2004; Ghezzi *et al.* 2014). Ethanol-related phenotypes, including alcohol tolerance and alcohol sensitivity, vary with genetic background (Morozova *et al.* 2015; Fochler *et al.* 2017; Morozova *et al.* 2018) and have allowed Drosophila to serve as a model for AUD and FASD, in addition to providing broader mechanistic insights on ethanol metabolism and its effects on neural tissue (McClure *et al.* 2011; Devineni and Heberlein 2013; Engel *et al.* 2019; Logan-Garbisch *et al.* 2014; Guevara *et al.* 2018; Belhorma *et al.* 2021).

Ethanol is often delivered to *D. melanogaster* via ethanol vapors or ethanol-supplemented food. Ethanol vapors can be administered during development in an ethanol bath, or even by placing ethanol on the cotton plug of vials. However, for behavioral quantification, ethanol is most commonly administered through an inebriometer, a vertical tube filled with ethanol vapors leading to loss of postural control, thereby quantifying total inebriation and or sedation time of groups of Drosophila (Cohan and Graf 1985; Weber 1988; Morozova *et al.* 2006). Capillary feeder (CAFÉ) assays can also be used to administer ethanol solutions for consumption to adult Drosophila, often providing them the choice of two or more solutions to calculate a preference index of the ethanol solution compared to an alternative solution (Ja *et al.* 2007). As individuals are given the choice of an ethanol solution, CAFÉ assays can be used to model AUD and quantify preference for/against ethanol-laced solutions. Supplementation of standard Drosophila food media with ethanol allows administration of ethanol throughout development, from embryo through adult, and can be used to model FASD while also measuring ethanol-mediated changes to Drosophila at specific developmental timepoints (McClure *et al.* 2011, Guevara *et al.* 2018; Morozova *et al.* 2018).

The protocols most commonly used to assess changes in behavior due to ethanol exposure are performed on groups of Drosophila. However, testing groups masks within-group phenotypic variation and ignores potential effects of social interactions between individuals on their behavior. Testing Drosophila individually prevents variation due to social interactions and increases statistical power with measurements on individuals. As in most fields, the effects of long-noncoding RNAs remains underappreciated – especially striking within ethanol research, as many long noncoding RNAs (lncRNAs) have known roles in stress-response. Recent research on Drosophila and ethanol has primarily been motivated by translational potential to humans, therefore important modulators of ethanol response with limited human implications, such as Drosophila-specific noncoding RNAs, remain understudied.

### *Drosophila models of Fetal Alcohol Spectrum Disorder*

In *D. melanogaster* models of FASD, developmental ethanol exposure is generally associated with decreased fitness. However, just as in other FASD model systems, genetic background influences these ethanol-mediated phenotypes. One GWA study examining viability and development time resulting from developmental ethanol exposure of 200 DGRP lines found genes enriched for development of neurons, the nervous system, and organs, as well as *Wnt* signaling and signal transduction (Morozova *et al.* 2018). This group also found *Cyclin E* to be a key node in a genetic interaction network, suggesting that cell cycle regulation, *Myc* activity, and ribosome biosynthesis are critical for response to developmental ethanol exposure (Morozova *et al.* 2018). Interestingly, *CycE* also has high expression in female ovaries, supporting the use of Drosophila to model FASD (Morozova *et al.* 2018).

Another mechanistic insight gleaned from Drosophila models of FASD is the involvement of insulin signaling. Insulin is a signaling factor upstream of mechanistic target of rapamycin (mTOR), both of which work to regulate gluconeogenesis and well as cell growth and differentiation. Ethanol inhibits the mTOR pathway and overactivation of *MTOR* rescued the effects of ethanol in zebrafish mutants without *MTOR* (Xu *et al.* 2013; Eberhart and Scott 2016). Although rat models have shown that signaling of insulin is reduced in the brains of FASD rat models (Xu *et al.* 2013), a Drosophila FASD model also found that overactivation of insulin signaling reduces ethanol-mediated effects (McClure *et al.* 2011). A subsequent study has identified separate roles for insulin signaling in ethanol-mediated changes to survival and developmental time and linked the drop in insulin signaling to increased fatty acid accumulation and oxidative stress, suggesting that the current knowledge of the role of insulin signaling in developmental ethanol exposure is incomplete (Logan-Garbisch *et al.* 2015).

Oxidative stress is another proposed mechanism for the ability of developmental ethanol exposure to alter development and behavior. Ethanol results in a two-pronged elevation in oxidative stress, as ethanol increases production of reactive oxygen species and decreases expression of genes involved in counteracting reactive oxygen species (Wu and Cederbaum 2003; Brocardo *et al.* 2011). Supporting this hypothesis, Drosophila larvae have increased expression of genes involved in oxidative stress response after exposure to alcohol (Logan-Garbisch *et al.* 2015, Belhorma *et al.* 2021). The transcriptional changes in oxidative stress response due to developmental ethanol exposure are associated with a neuroprotective effect against aging, a consequence of oxidative stress (Belhorma *et al.* 2021). Aged Drosophila with a history of developmental ethanol exposure do not experience the same degree of age-related decline in performance on negative geotaxis assays, measures of central nervous system function, compared to their control counterparts (Belhorma *et al.* 2021).

*D. melanogaster* models of FASD are particularly intriguing because ethanol is a naturally occurring component of the *D. melanogaster* environment. Drosophila lay their eggs in and eat microorganisms on rotting fruit, an environment relatively high in ethanol. Guevara *et al.* (2018) found that developmental ethanol exposure in a *D. melanogaster* model of FASD was associated with decreased food consumption during development as well as upregulation of *neuropeptide F*, which is important in feeding behavior, and provided a protective effect against ethanol-mediated lethality during development. While the decreased food intake and transcriptomic changes are important to consider in future Drosophila FASD studies, these data also suggest that *neuropeptide F* expression may be a *D. melanogaster*-specific adaptation to counteract developmental environments with high ethanol levels.

Ethanol exposure may also lead to epigenetic changes. Adult Drosophila with repeated exposures to ethanol result in progeny with decreased sensitivity to ethanol, even when at least one week had elapsed since last ethanol exposure (Bonilla *et al.* 2021). However, the progeny did not necessarily develop functional tolerance (Bonilla *et al.* 2021). Histone methyltransferase G9a is a H3K9 epigenetic regulator responsible for regulating insulin-like peptide genes (Shimaji *et al.* 2017) and the histone deacetylase *Sirt1* is downregulated and leads to changes in gene expression itself after ethanol exposure (Morozova *et al.* 2006; Engel *et al.* 2016). *Sirt1* is also associated with insulin signaling in *D. melanogaster*, further supporting the role of insulin in ethanol response, and the connection between aging, ethanol, oxidative stress, and insulin. Early developmental ethanol exposure in mammalian models has led to evidence of long-term changes in epigenetic reprogramming, methylation, and histone modifications (Kleiber *et al.* 2014).

Some Drosophila models of FASD examining transcriptional changes resulting from developmental ethanol exposure describe different observations. For example, one manuscript focuses on the importance of insulin in Drosophila models of FASD (McClure *et al.* 2011), but another manuscript does not corroborate this finding (Morozova *et al.* 2018). These differences may be due to experimental design, may be artifacts, or may be genotype-specific adaptations to ethanol in the Drosophila natural environment; additional data are necessary to attempt to resolve these discrepancies. Small scale studies indicate epigenetics, transgenerational inheritance, and aging are important in response to ethanol (McClure et al 2011, Logan-Garbisch *et al.* 2015, Belhorma *et al.* 2021, Bonilla *et al.* 2021, but very few larger scale studies (Morozova *et al.* 2018) have been conducted to verify and expand upon these findings, especially across multiple genetic backgrounds and experimental setups. Additional insights on the effects of ethanol on Drosophila will come as new epigenetic, single-cell, and multi-omic technologies become mainstream.

### *Role of Sex*

Sex is an important biological concept that should be accounted for in research studies, especially those on an individual's response to ethanol. AUD disproportionately affects more males than females (Substance Abuse and Mental Health Services Administration 2020), and the physiological response to ethanol differs across sexes (Hasin and Grant 2015). Alcohol intake of the mother carrying the fetus is of primary concern in FASD, although alcohol intake can also affect sperm quality in males (Finelli *et al.* 2022). Examination of a single sex in any model system ignores the possibility of sex-specific changes. There is strong evidence in *D. melanogaster* that most traits differ across sexes, including prenatal ethanol exposure, sleep, brain morphology, and startle-induced locomotor response (Mackay and Huang 2018). However, some traits in humans do not appear to be sex-specific or lack the quantity of data needed to analyze sexes separately, as is the case with some rare disease phenotypes. In these cases, assessment of both sexes in model systems and then examining traits that are similar in both sexes or conducting a pooled analysis across both sexes is more likely to be a true representation of the disease state in humans, as opposed to assessing a single sex alone. The degree to which sexual dimorphism in Drosophila models of rare disease impacts translational potential to humans remains understudied.

# *Drosophila melanogaster* as a model for rare neurodevelopmental disorders of chromatin modification

Although a disease must affect fewer than one in 200,000 individuals to be classified as rare, rare diseases as a whole affect over 30 million people, or about 1 in 10 people, in the United States alone (National Human Genome Research Institute 2020). Most of these diseases have small patient populations, which makes diagnosis, advocacy, characterization, and scientific study of the disease challenging. Environmental and genetic factors are responsible for about 25% and 75% of rare diseases, respectively (Global Genes 2023). Despite the high fraction of rare diseases with genetic causes, only about 5% of rare diseases have treatments and less than half of patients with a rare genetic disease receive a molecular diagnosis (Marwaha *et al.* 2022; Global Genes 2023). However, government-backed incentivization of pharmaceutical development for and research on rare disorders has been recently increasing (National Human Genome Research Institute 2020). Children are disproportionately affected by rare diseases, as half of rare disease patients are children, and long diagnostic odysseys may lead to misdiagnosis or a lack of a diagnosis in younger individuals (Marwaha *et al.* 2022; Global Genes 2023).

## *Drosophila as a model for rare disorders*

Genetic study of rare human disorders is extremely challenging in humans and higher-order model organisms, as analyses often require greater levels of statistical power than can be attained given ethical, temporal, and monetary restrictions on scientific research. Drosophila are well suited for the study of rare genetic disease, as balancer chromosomes allow for maintenance of otherwise deleterious variants and large numbers of Drosophila can be reared at low cost. Genome manipulation to humanize Drosophila by inserting human wildtype and mutant cDNA into the Drosophila genome can help clinical teams assess the putative function of a specific variant of unknown significance and provide evidence as to whether the variant may be disease-related in a human patient (Ugur *et al.* 2016; Link and Bellen 2020). Forward genetic screens in Drosophila have identified novel genes associated with human disease (Bell *et al.* 2009; Yamamoto *et al.* 2014; Link and Bellen 2020) and can identify pathways involved in disease pathogenesis that can be targeted for pharmaceutical intervention. Drosophila can also be used as an economical, low-risk, and high-throughput method to assess the effectiveness of possible pharmaceutical interventions for individual patients (Bell *et al.* 2009, Hope *et al.* 2022). In addition to assessing the impact of a single variant, reference panels such as the DGRP allow for study of the effect of a variant across multiple genetic backgrounds on disease phenotype (Talsness *et al.*2020; Hope *et al.* 2022). Although behavioral and whole-organismal phenotypes can only be performed in Drosophila adults and larvae, Drosophila cell culture remains a valuable tool to gain mechanistic effects of genetic variants.

The conservation of neurological development between Drosophila and humans makes Drosophila particularly useful in studying disorders with neurological presentations such as seizures, neurodegeneration, developmental delay, and intellectual disability (Coll-Tane *et al.* 2019; Link and Bellen 2020). As examples, work on rare diseases in Drosophila has led to diagnosis of human patients with microcephaly (Cavallin *et al.* 2017), identification of therapeutic targets of NLGY1-deficiency (Hope *et al.* 2022) and strengthened a novel association of a RhoGTPase with a developmental and epileptic encephalopathy (Straub *et al.* 2018). Drosophila was also instrumental in mechanistic findings for Kleefstra syndrome, which is caused by variants in the chromatin remodeler *EHMT1*, as studies demonstrated that variants in the Drosophila ortholog of *EHMT1* (*G9a*) resulted in changes to learning and memory behaviors and gene expression (Kramer *et al.*2011). Some patients with Kleefstra phenotypes did not harbor variants in *EHMT1*. Drosophila was then used to test for genetic interactions between *EHMT1* and genes of interest that harbored variants possessed by patients with wildtype *EHMT1* (Kleefstra *et al.* 2012). This led to diagnosis for patients as well as increased basic science knowledge of the *EHMT1* chromatin remodeling complex.

Although Drosophila are established tools for modeling rare neurodevelopmental disorders, many studies do not use available technologies to reach their full potential. Increased accessibility of RNAseq allows researchers to examine an entire affected animal, while simultaneously assessing transcriptional profiles of affected and unaffected tissues. Tissue-specific studies could increase the utility of Drosophila to model disorders with genetic underpinnings historically considered outside of neurological development or discover non-neurological manifestations in Drosophila models of neurodevelopmental disorders. Single-cell multi-omics would provide additional levels of detailed information on the effects of ethanol on different cell types. Another area in which many current Drosophila models fall short is regarding the effects of genetic background, as common practice is to use the Drosophila strain with maximal convenience to the researcher, skewing comparisons across studies.

Drosophila could also be useful in disentangling the effects of environment on neurodevelopmental disorders. Human patients are not exempt from environmental factors, such as developmental alcohol or heavy metal exposure, which can muddy the clinical picture. Assessment of a disorder across multiple environments remains a largely untapped area of rare disease research. Recently, findings from some Drosophila models of a rare congenital disorder of glycosylation have been immediately translatable to humans (Hope *et al.* 2022; Perlara PBC 2023). Additional successes will indicate whether immediately translatable studies are coincidental or an indication of what is to come as the field advances.

### *Chromatin modification and disease*

In addition to Kleefstra syndrome, SWI/SNF-related intellectual disability disorders (SSRIDDs) and Cornelia de Lange syndrome (CdLS) are rare neurodevelopmental disorders that stem from mutations in chromatin remodeling protein complexes. Chromatin is the collection of nuclear DNA, associated proteins, and any associated modifications (e.g. methyl groups). Chromatin state is tied to the accessibility of the

DNA, where open chromatin can be bound by molecular machinery responsible for transcription and or replication (Boltsis *et al.* 2021). In its compact form, DNA is normally wound around proteins called histones. Histones, as well as the DNA itself, can be modified (*e.g.* with methyl, acetyl and phosphoryl groups) to change the strength of the bond between negatively charged DNA and positively charged histones, thereby modulating the accessibility of the DNA. Protein complexes rely upon chromatin accessibility to facilitate long-range interactions and loading of transcriptional machinery, which provide feedback and lead to other changes in chromatin state (Boltsis *et al.*2021). These changes in DNA accessibility lead to changes in transcription and thereby gene expression throughout the genome, potentially leading to widespread downstream effects and disease development.

The role of chromatin modification in disease is becoming increasingly complicated. The plethora of downstream effects stemming from aberrant chromatin modifications makes identification of mechanisms for diseases related to chromatin modification challenging. Some chromatin modifying enzymes have significant roles outside of chromatin modification, while other chromatin modifying complexes change in composition across tissue and time  or may have individual complex subunits with roles unrelated to chromatin, further obscuring potential disease mechanisms. New disease-gene associations are increasing the number of chromatin-related genetic disorders, and epigenetic signatures have failed to be detected in individuals with pathogenic variants in genes related to chromatin modification (Nixon *et al.* 2019; Barish *et al.* 2020; Aref-

Eshghi *et al.* 2021). Phenotypically, pathogenic variants within members of the same chromatin modification complex result in disparate clinical findings, suggesting the presence of genetic modifiers. The full picture of chromatin modification in disease is far from clear.

### *SWI/SNF-related intellectual disability disorders*

The mammalian SWI/SNF (mSWI/SNF) or Brahma-Related Gene 1 (BRG1)/ Brahma (BRM) associated factor (BAF) complex is a large adenosine triphosphate (ATP)-dependent chromatin remodeler that functions as a reader by recognizing histone modifications, binding histone-wrapped DNA in a largely non-sequence-specific manner through zinc-finger and bromo domains (He *et al.* 2020). The BAF complex is antagonistic to Polycomb repressive complex (PRC)-mediated chromatin silencing through competitive binding and removal of PRC1 and PRC2, and/or ejection of the histone leading to free DNA (Kadoch *et al.* 2017; He *et al.* 2020). The BAF complex, therefore, is primarily associated with expression activation, especially for tissue-specific expression and subsequent cell differentiation (Centore *et al.* 2020). The mSWI/SNF complex has three primary forms, the canonical, polybromo-associated, and non-canonical BAF complexes, although there are 29 possible components, of which up to 15 are incorporated into BAF complexes at any given point throughout development (Centore *et al.* 2020; He *et al.* 2020). Some of the more critical components include SMARCA2/4, which are BAF ATPases, and SMARCB1, which facilitates proper positioning of and interacts with the DNA-bound nucleosome, while most other subunits, including ARID1A/B, and SMARCC1/2, serve as a core and structural scaffold (He *et al.* 2020). In addition to transcriptional regulation and tissue differentiation, the mSWI/SNF complex also has roles in double-strand break repair (Cenik and Shilatifard 2021). Defects in components of the mSWI/SNF complex are associated with cancer, autism spectrum disorder, schizophrenia, Kleefstra's syndrome spectrum, and SSRIDDs including Coffin-Siris and Nicolaides-Baraitser syndromes (Sokpor *et al.* 2017; Cenik and Shilatifard 2021).

SSRIDDs are a collection of autosomal dominant neurodevelopmental disorders that arise from variants in the human SWI/SNF complex (Bogershausen and Wollnik 2018). There are no distinct diagnostic criteria or treatment options for SSRIDD patients, in part due to the wide phenotypic spectrum. Classic Coffin-Siris syndrome (CSS) is characterized with developmental delay, intellectual disability, fifth-digit abnormalities, hypotonia and abnormal hair growth (Coffin and Siris 1970; Santen *et al.* 2013). However, under the larger umbrella of SSRIDDs, patients may also or instead present with seizures, Dandy-Walker malformations and agenesis of the corpus callosum, course facial features, hearing and eye abnormalities, and cardiac and renal malformations (Schrier Vergano *et al.* 2021). Patients with variants in *ARID1B* are the most commonly reported SSRIDDs patients and typically present with a milder phenotype, whereas patients with variants in *SMARCB1* present with a more severe clinical picture (Bogershausen and Wollnik 2018; Schrier Vergano *et al.* 2021). Interestingly, some patients with *ARID1B* variants may only present with mild intellectual disability or autism spectrum disorder, and no other notable findings. Nicolaides-Baraitser syndrome is an additional form of SSRIDD but is only associated with variants in *SMARCA2* (Nicolaides and Baraitser 1993; Bogershausen and Wollnik 2018). Sufficient prognostic data on SSRIDDs has not yet been collected to assess longer term patient outcomes.

The exact mechanism of SSRIDDs is largely unknown. SSRIDDs are associated with loss of function, gain of function, and dominant negative mutations, whether missense, nonsense, or deletion variants (Bogershausen and Wollnik 2018). SSRIDDs have a wide phenotypic spectrum and the mSWI/SNF complex has a wide range of roles, including those in transcriptional regulation, DNA repair, maintenance of pluripotency, and cellular differentiation (Cenik and Shilatifard 2021). Furthermore, the exact composition of the BAF complex varies across tissues and developmental time points, and some gene families within the mSWI/SNF complex (e.g. *ARID* genes) appear to be less or more frequently mutated than others (Kadoch and Crabtree 2015; Bogershausen and Wollnik 2018). Interestingly, although genes associated with SSRIDDs are frequently mutated in cancers, cancers have only been reported in three individuals with Coffin-Siris syndrome thus far, each of whom had a variant in a different mSWI/SNF component (Schrier Vergano *et al.* 2021). However, pathogenic *SMARCB1* variants can cause schwannomatosis and rhabdoid tumor predisposition syndrome, and *SMARCE1* variants have been identified in patients with spinal and clear cell meningiomas, though these patients do not have other SSRIDD-related phenotypes (Kadoch and Crabtree 2015; Schrier Vergano *et al.* 2021). Thus, further study is needed on the cancer risk for

SSRIDD patients and suggests the potentially subunit- and disease-specific mechanisms for pathogenesis of SSRIDD and SWI/SNF-related tumorigenesis.

Studies in mice and Drosophila have shown that lack of specific SWI/SNF components, including *BRG1*, *BAF47*, and *BAF155*, is associated with early embryonic lethality; and gene knockdown studies for *BRG1* showed decrease pluripotency and cell proliferation as well as widespread changes in transcription (Alfert *et al.* 2019). The non-canonical BAF complex specifically binds to CTCF and topologically associating domains and is therefore thought to play a role in long-range chromatin organization (Michel *et al.* 2018), providing yet another potential mechanism for SWI/SNF-related disease pathogenesis. Heterozygous changes in the neural progenitor BAF and neural BAF complexes lead to neural tube defects and changes in the cerebellum, cortex and midbrain of mice (Sokpor *et al.* 2017; Alfert *et al.* 2019). BAF complexes are also important in development of glia and neurons, as well as cardiac and muscle development (Sokpor *et al.* 2017). Thus, the exact mechanisms of SSRIDDs are multifaceted and likely variant-specific, complicating functional studies on SSRIDDs.

Despite the plethora of research on the SWI/SNF complex and orthologous complexes, the role of the mSWI/SNF complex in SSRIDD pathogenesis remains unknown. The lethality associated with loss of mSWI/SNF complex members, the widespread impact of mSWI/SNF-mediated chromatin modification, and the involvement of mSWI/SNF complex members in roles outside of chromatin modification are current obstacles in discovering SSRIDD mechanisms. Insights on mechanism would support or refute the ongoing shift in classification from individual disorder to a spectrum-based classification and the use of an umbrella term SSRIDD, which are currently based largely on the varying clinical pictures associated with variants in mSWI/SNF complex components, not molecular or mechanistic evidence. Molecular evidence could also help answer the question of why cancer has been observed in some SSRIDD patients, but not others, given the high rates of mSWI/SNF-related variants in tumors.

Even as new SSRIDD-related genes are discovered, we still cannot explain the degree of clinical variability for patients with pathogenic variants in the same gene, or for patients with pathogenic variants in different mSWI/SNF-related genes. Hypotheses for why and how this variation exists in SSRIDDs have not yet been tested. Clinically, long-term prognosis and cancer rates of SSRIDD patients is unknown, partially due to the small number of known SSRIDD patients.

### *Cornelia de Lange syndrome*

Although the cohesin complex is named for its role in sister chromatid cohesion during the G1 and S phases of cell division, this highly conserved, ring-shaped complex also facilitates transcriptional regulation, DNA repair, and long-range interactions across the genome, though less is known about these roles (Nasmyth and Haering 2009). Cohesin has four primary protein components: SMC1, SMC3, RAD21 and Stromalin (SA1 or SA2). SMC1 and SMC3 make up the the cohesin ring along with RAD21 and SA1/SA2 binds to RAD21 to facilitate interactions between multiple cohesion rings (Nasmyth and Haering 2009; Zhu and Wang 2019). Accessory proteins, such as NIPBL, MAU2, and HDAC8 are involved in loading, recruitment, and stabilization of cohesin onto chromatin, while the accessory protein CTCF is important for binding at transcriptional start sites and proper gene expression (Nasmyth and Haering 2009; Deardorff *et al.* 2012; Zhu and Wang 2019). Defects in cohesin can lead to aneuploidy, increased apoptosis, and decreased cell division, as sister chromatids are not held together properly during cell division, as well as altered gene expression, as long-range chromatin and promoter-enhancer interactions and proper localization of cohesin at transcriptional start sites become disrupted (Zhu and Wang 2019).

Variants in *NIPBL*, *SMC1A*, *SMC3*, *RAD21*, *BRD4*, and *HDCA8* are associated with Cornelia de Lange syndrome (CdLS) (Deardorff *et al.* 2020). CdLS patients have distinct facial gestalt, including synophrys, as well as intellectual disability, digit anomalies, and growth retardation (Kline *et al.* 2018). Most CdLS is autosomal dominant, but *HDAC8*- and *SMC1A*- CdLS are X-linked (Deardorff *et al.* 2020). The severity of the phenotype is often, but not always, dependent on the type of variant and the gene containing the variant. Because of this phenotypic variation, some patients present with clinical features resembling another disorder: patients with *SMC1A* variants present with a phenotype resembling the X-linked intellectual disability disorder Rett syndrome, or without many of the hallmark CdLS features (Kline *et al.* 2018). Due to the small number of CdLS cases and the relatively large number of mosaic CdLS patients, genotype-phenotype correlations and predicted patient outcomes are weak or non-existent, depending upon the genotype (Deardorff *et al.* 2020). Most CdLS patients will survive into adulthood, though the most common causes of death in infants and children are congenital diaphragmatic hernia, and respiratory issues, respectively (Schrier *et al.* 2011). Variants in *RAD21* and *SMC1A* can also be associated with autosomal recessive visceral neuromyopathy (Mungan syndrome) and early-infantile epileptic encephalopathy type 85, respectively (Deardorff *et al.* 2020).

In addition to variants in structural components of cohesin and *NIPBL* that give rise to classic CdLS, variants in other cohesin-associated genes have been associated with CdLS-like phenotypes, suggesting the presence of a CdLS spectrum under the larger umbrella term of cohesinopathy (Kline *et al.* 2018). The mechanism of cohesinopathies as a group remains in question. Variants in *ESCO2*, a gene responsible for cohesion establishment in S phase (Alomer *et al.* 2017), are associated with the autosomal recessive ESCO2-Spectrum disorder, previously known as Roberts-SC phocomelia syndrome (Goh *et al.* 2010). These patients have severely shortened long bones, intellectual disability, decreased growth, and craniofacial, renal and cardiac anomalies due to increased apoptosis and reduced cell proliferation (Goh *et al.* 2010). Interestingly, although genes associated with both CdLS and ESCO2-Spectrum disorder are related to regulation or structure of the cohesin complex, cells from patients with variants in *SMC1A*, *SMC3*, *NIPBL*, *HDAC8*, and *RAD21* do not show issues with chromatid cohesion and instead show only altered gene expression and may have changes in DNA repair (Dorsett and Krantz 2009).

CdLS can be loss of function (for variants in *BRD4*, *HDAC8*, *NIPBL* and *RAD21*) or gain of function/dominant negative (for most variants in *SMC1A* and *SMC3*) (Deardorff *et al.* 2020). The exact mechanism for CdLS pathogenesis has not been resolved, but clues from Drosophila provide a starting point. Lack of SMC1, SA and RAD21 led to impaired axon pruning in Drosophila (Pauli *et al.* 2008, Schuldiner *et al.* 2008), thereby indicating that cohesin is critical for proper development of the nervous system. Another study in Drosophila observed Nipped-B and cohesin throughout chromatin, bound to hundreds of active genes, especially those regions bound by RNA polymerase II (Misulovin *et al.* 2008), suggesting that cohesin and Nipped-B are instrumental in gene expression. The lack of impaired chromatid cohesion in CdLS patients and the lack of known mutations in critical cohesin subunit domains suggests that CdLS pathogenesis does not arise from reduced cohesin activity, but instead is due to an alteration in cohesin function. For example, changes in the distribution and rate of cohesin and or NIPBL binding to the genome likely leads to changes in gene expression that affect key developmental pathways (Dorsett and Kratz 2009).

Many of the same outstanding questions for SSRIDDs remain for CdLS. The basic mechanisms of disease pathogenesis, regardless of the subunit or cofactor, remain unknown. Work in Drosophila has indicated that CdLS may arise through errant function of a secondary role of the cohesin complex, but exactly how and when the cohesin complex regulates transcription, or how these changes in transcriptional regulation result in CdLS remains unknown. Mechanistic studies are challenging for CdLS-associated genes, as loss of these genes is incompatible with life. Furthermore, it is not known why clinical presentations vary within and across CdLS-associated genes, or why some CdLS-associated genes also harbor pathogenic variants associated with other disorders. Mechanistic insight would also help determine the accuracy of a spectrum-based approach to CdLS. CdLS has been partially characterized using the Drosophila model, but hypotheses beyond characterization have not yet been tested in animal models. Despite larger numbers of CdLS patients, long-term prognosis and best practices in clinical management is still unknown.

### *Cancer*

Given the integral nature of chromatin state to the cell cycle, DNA replication, and transcription, it is not surprising that many cancers possess mutations in genes tied to chromatin modification. These variants are in chromatin readers, writers, and erasers, as well as genes responsible for catabolism and anabolism of chromatin modification substrates (Zhao *et al.* 2021). Additionally, mis-regulation of chromatin modifications are just as frequent causes of cancer as specific genetic variants (Marine *et al.* 2020). One of the more common genetic causes of cancer is due to variants in components of the mammalian SWI/SNF (BAF) complex, which are estimated to exist in greater than 20% of all cancers and lead to both loss-of-function and gain-of-function cancers (Kadoch and

Crabtree 2015). *ARID1A* encodes the most commonly mutated BAF subunit in cancer, and variants in *ARID1A* are found in multiple tumor types (Kadoch and Crabtree 2015). Many genes encoding cohesin complex members and associated proteins are also mutated in cancer, such as *NIPBL* and *SMC1A*, which are mutated in about 4.9 and 1.7 percent of tumor samples, respectively (COSMIC database: cancer.sanger.ac.uk; Tate *et al.* 2019). Dysfunctional cohesin could lead to cancer through its roles in chromatid cohesion, DNA repair, or transcriptional regulation. Aneuploidy as a result of cohesin dysfunction can serve as the "second hit" in the two-hit hypothesis, whereas polyploidy can lead to overexpression of a proto-oncogene (Di Nardo *et al.* 2022). Histone deacetylases, including *HDAC8*, are also associated with cancer development through alteration of transcription as a result of deacetylated histones and other proteins (Li and Seto 2016; Zhao *et al.* 2021).

Cancer is a major area of biological research with too many yet unanswered questions to be discussed here in full. However, there are a subset of unanswered questions related to disorders of chromatin modification, namely why some subunits of chromatin modification protein complexes are more associated with cancer risk and/or are detected more often in tumors of otherwise healthy individuals. Unfortunately, given the rarity of patients with disorders of chromatin modification, these questions will have to be addressed from a cancer perspective, as opposed to an SSRIDD- or CdLS-based approach.

### *Drosophila as a model for SSRIDDs and CdLS*

Using Drosophila to model SSRIDDs and CdLS provides specific advantages, including easy generation of a large number of afflicted individuals, a high degree of human-Drosophila conservation for SWI/SNF and cohesin complexes, RNAi lines that mimic the heterozygous nature of autosomal-dominant disorders, and balancer chromosomes that facilitate study of otherwise lethal alleles. Although craniofacial, digit, and hair anomalies are challenging to model in Drosophila, neurological phenotypes can more easily be assessed. For example, while sleep issues, seizures, and structural brain anomalies are not diagnostic for CdLS or SSRIDDs, these phenotypes are commonly observed in CdLS and SSRIDD patients (Kline *et al.* 2018; Schrier Vergano 2021) and therefore could be considered representative phenotypes for Drosophila models of CdLS and SSRIDD. Drosophila has been used to model other common SSRIDD and CdLS phenotypes, including low muscle tone and intellectual disability (Bellen *et al.* 2019; Coll-Tane *et al.* 2019).

*D. melanogaster* is particularly well suited to model CdLS, as much of the mechanistic work on CdLS pathogenesis performed thus far was performed in Drosophila cell lines (Dorsett 2016), and a previous model of *NIPBL*-associated CdLS has been well characterized in *D. melanogaster* (Wu *et al.* 2015). A previous study also used RNAi-based SSRIDD Drosophila models to look for specific changes in brain morphology and neuronal development (Chubak *et al.* 2019). Although there is no DNA methylation in Drosophila, neither the human SWI/SNF nor cohesin complexes rely upon DNA methylation as the sole disease mechanism. Histone-based chromatin modifications and changes in chromatin due to long-range interactions are still conserved from Drosophila to humans.

Variants associated with other neurodevelopmental disorders are not uncommon in CdLS and SSRIDD patients. For example, variants in *KMT2A*, a histone methyltransferase, and in members of the human chromatin remodeler SWI/SNF complex, have been identified in patients that met clinical diagnostic criteria for classic CdLS (Parenti and Kaiser 2021). Furthermore, variants in genes with roles perhaps more distantly related to cohesin, such as *DDX23*, an RNA helicase, and *NAA50*, a component of the NatE complex, which is involved in post-translational protein modification, have also been identified in patients with clinical presentations similar to or meeting the diagnostic criteria for CdLS (Parenti and Kaiser 2021). Currently, there is debate in the literature as to classification of SSRIDDs, such as whether variants in *SOX11*, *BICRA* and *SMARCD1* should be associated with a form of Coffin-Siris syndrome (Schrier Vergano *et al.* 2021; Al-Jawahiri *et al.* 2022), or whether CSS and NCBRS or other clinical differentials are part of the same disease spectrum (Wieczorek *et al.* 2013; Aref-Eshghi *et al.* 2018; Bogershausen and Wollnik 2018). Use of the umbrella term SSRIDD allows inclusion of genetic disorders together that were previously considered distinct from one another. Still, individuals with variants in *PHF6* were initially diagnosed with CSS based on clinical findings, despite a molecular diagnosis of Borjeson-Forssman-Lehmann syndrome (Wieczorek *et al.* 2013).

There are a few possibilities for the abundance of overlap across supposedly distinct disorders, none of which are mutually exclusive. These disorders could all be mechanistically linked due to the wide range of roles that the cohesin and mammalian SWI/SNF complexes may have in the cell. The discrepancy could be a result of the clinical findings, as the diagnostic criteria for CdLS or SSRIDDs may not be specific enough and/or the phenotypic spectrums for CdLS and SSRIDDs are too wide to be captured well with discrete diagnostic criteria. A third possibility is that there are secondary variants in the genetic backgrounds of some patients that could contribute to variation in disease presentation across disorders that may have overlapping phenotypes, whether molecular or clinical (Rahit and Tarailo-Graovac 2020). These secondary variants and their associated genes are termed "genetic modifiers." Studies have not yet investigated the possibilities of genetic modifiers, nor have resolved the mechanisms for these disorders, in humans or in Drosophila.

Studies on SSRIDDs and CdLS in Drosophila (Pauli *et al.* 2008, Schuldiner *et al.* 2008, Misulovin *et al.* 2008, Pauli *et al.* 2010, Wu *et al.* 2015, Chubak *et al.* 2019) have primarily consisted of experiments involving a single tissue- or cell- type. It is not known whether the phenotypes assessed are accurate representations of phenotypic changes in human SSRIDD and CdLS patients. Only one study has modeled either SSRIDDs or CdLS and considered more than one protein complex component simultaneously, in the same genetic background (Chubak *et al.* 2019). Further studies on how each subunit contributes to disease individually, or on the degree to which alterations in each complex member or cofactor cause similar or dissimilar changes to the whole organism are needed. Transcriptomic profiles performed on whole-fly CdLS models were also restricted to a specific tissue--type (Wu *et al.* 2015), though changes in epigenetic profiles, protein levels, and tumor development have not been quantified on any level in any models.

### *Genetic modifiers – an overview*

Over the past few decades, there have been an increased number of studies of genetic modifiers of rare and common diseases.Traditionally, genetic modifiers are defined as a genetic variant that modulates the phenotype of another genetic perturbation. Phenotypic modulation may occur through epistasis (gene-gene interaction) where the presence or absence of the epistatic partner leads to a change, in the magnitude or direction, of the phenotypic trait stemming from variation at the focal gene (Mackay 2014). Epistasis can suppress or enhance the phenotype of the focal allele and the the magnitude of epistasis depends on allele frequencies in the population (Mackay 2014). Genetic modifiers were identified as early as 1914 (Dexter 1914) in Drosophila, where genetic background accounted for non-Mendelian ratios of *truncate* wing mutants. Although linkage mapping was used to identify genetic modifiers for most of the 20th century, genetic modifiers can also be identified through association mapping (GWAS studies). GWA studies examine the association between a given genomic locus (*e.g.* SNP) and the phenotype of interest across all individuals in the study, resulting in a prediction in the relationship between the locus and the phenotype. The locus with the most significant association in a genomic region is not necessarily the causative allele, although the associated locus is in linkage disequilibrium (LD), or segregates with, the causative allele. The type of allele, computational prediction and functional annotation of polymorphisms that were not genotyped for the GWAS but are in LD with the associated allele can lead to identification of a presumed causative, or in the case of genetic modifiers, modifying allele (Tam *et al.* 2019). Once identified, genetic modifiers may explain phenotypic variation observed across patients and provide basic insight as to the disease mechanism.

GWA results depend on the study design. The minimal effect size that can be detected is determined by the sample size: detection of smaller effect sizes requires more individuals (Tam *et al.* 2019). Population stratification can lead to false positive associations and must be controlled. There is a trade-off between larger numbers of genomic loci (more dense genotyping) and the multiple testing correction, which is based on the number of independent association tests performed and is used to limit false positives (Mackay 2014). Identification of genetic modifiers for rare diseases is challenging in human populations. To achieve statistically significant results, GWAS require large numbers of unrelated individuals, which do not exist for most rare diseases. The need for large studies is especially true for quantitative traits and common diseases that arise from multiple small-effect alleles. Rare diseases, or diseases that arise due to alleles with larger effect sizes do not require such large population sizes, although by definition, there are not enough individuals with a rare disease to constitute even a “small” population. GWA-based methods for identification of genetic modifiers are also not feasible, a barrier to the growing field of precision medicine. Accurate molecular and clinical phenotyping are additional barriers for GWA analyses of rare disorders. Alternative methods for identification of genetic modifiers for rare disorders using only human data would be beneficial to human genetics but remain unproved. Animal models, including rodents, Drosophila and worms, are therefore valuable in the identification of genetic modifiers, especially for rare disorders, as large numbers of unrelated individuals can be reared and phenotypically quantified with minimal environmental variance.

### *Genetic modifiers and rare disease*

Further contributing to the complications of identifying genetic modifiers in humans is that not all secondary variants reported may be related to pathogenesis. Unfortunately, identification of such instances is rare and demonstrating the resulting molecular impact is extremely challenging. A patient may present with two or more disease-causing variants, each of which is associated with a separate, unrelated genetic disorder. As mentioned earlier, this may be the case for some patients with SSRIDDs and/or CdLS. The subsequent clinical phenotype in such cases is often difficult to disentangle and may blur straightforward diagnosis of either disorder. These two variants (or the molecular phenotypes stemming from these variants) may interact with one another to modulate disease presentation (i.e., act epistatically), and therefore be considered genetic modifiers of one another (Rahit and Tarailo-Graovac 2020). However, they might be genetic modifiers through perhaps a non-traditional or unanticipated mechanism, as studies typically identify genetic modifiers based on the ability of the specific variant to change disease presentation in the presence of only a single disorder, not the extent to which known disease-causing variants are codominant in the presence of other known disease-causing variants (Rahit and Tarailo-Graovac 2020). However, development of software to predict the impact of two variants in the same individual would allow laboratory geneticists to begin to disentangle these effects in human populations, yet these have not yet been introduced. High throughput animal models could also facilitate prediction of genetic interactions between potentially pathogenic variants. Traditionally, identification of and impact from genetic modifiers of rare diseases has been ignored; other than the somewhat recent idea that such modifiers may exist for rare disorders, there is little known about their frequency, magnitude of effect, or how to best identify them. Outside of human populations, identification of genetic modifiers for rare disorders using model organisms shows the most promise.

### *Identification of genetic modifiers in Drosophila*

*D. melanogaster* has been used to identify genetic modifiers for over a century. Recent efforts to identify genetic modifiers for human disorders in Drosophila have been aided by widely available CRISPR, RNAi knockdown and RNAi overexpression lines which allow for relatively easy modeling of human disorders, and the DGRP, which is a large repository of well-characterized unrelated genetic backgrounds. Screens to identify modifiers have also been performed in cell culture using *D. melanogaster* cell lines (Luhur *et al.* 2019), though these are perhaps not as indicative of a macro-scale phenotype (such as sleep/movement or neurodegeneration) and fall prey to the same pitfalls of all cell culture experiments. Screens to identify candidate modifiers in cell culture may screen thousands of induced mutations in the presence of a drug, for example, that inhibits the gene of interest and then phenotype the cell for gene expression, cell death, etc. and examine the associated cell genotype (Dalton *et al.* 2022).

Screens to identify candidate genetic modifiers for human disorders in living Drosophila typically first induce a mutation or genetic perturbation relevant to the disease phenotype in a single genetic background, then cross that line to as many genetic backgrounds as possible (e.g. the DGRP, lines with RNAi-mediated perturbations of candidate modifiers) and score resulting F1 progeny for the phenotype of interest (Cukier *et al.* 2008, Lavoy *et al.* 2018, Talsness *et al.* 2020, Dalton *et al.* 2022). Depending on the experimental design, the phenotypic and genotype data can be used to perform a GWA study. Functional validation of associated loci can be assessed by quantifying other perturbations in the associated gene (*e.g.*, induced mutations, naturally occurring variants, or changes in expression due to RNAi) on the phenotype of interest, or by quantifying changes in phenotype as a result of a genetic perturbation in the candidate modifier within the genetic background containing the original focal variant. The latter method provides stronger functional evidence for epistasis between the candidate and the focal gene, though is much harder to accomplish while still preserving genetic background. Relevance of candidate modifiers for human disease (as opposed to Drosophila-specific effects) can be at least partially quantified through databases of functional annotation, such as Gene Ontology, and prediction of gene conservation from Drosophila to humans. Despite the pervasive nature of epistasis (Mackay 2014)and evidence for variable phenotypes in human disorders, relatively few studies have examined the role of genetic modifiers, especially in rare disease pathogenesis. This is likely due to the lack of statistical power in these populations, although this is no longer a barrier when using Drosophila, with tools such as the DGRP and tissue-specific drivers available.

Limiting expression of a specific deleterious genotype to a tissue that is not critical for life, such as the Drosophila eye or wing, allows study of genotypes that would otherwise sterilize and or kill the animal. Genetic modifier screens in Drosophila are only limited by the number of genetic backgrounds available and the ease of phenotyping. It should be noted that genetic modifiers can be identified for genetic perturbations of large or small effect sizes, though to date fewer studies have identified modifiers of small effect size. Genetic screens in living Drosophila have been used to identify modifiers for Drosophila phenotypes of apoptosis (Palu *et al.* 2019), head morphology (Özsoy *et al.* 2021), and olfaction (He *et al.* 2016), as well as candidate modifiers for human disorders DPAGT1-CDG (Dalton *et al.* 2022), Parkinson's disease (Lavoy *et al.* 2018), and NGLY1 deficiency (Talsness *et al.* 2020).

Despite recent successes in identifying candidate genetic modifiers for human diseases using the Drosophila model, candidate modifiers have not been widely validated in human populations. Thus, the translatability to humans of these methods remains in question. It has also not been tested whether the same genetic modifiers can be identified using different approaches in the same experimental conditions, *e.g.*, does a screen using a wing-tissue phenotype give the same results as one using an eye-tissue phenotype?

Given the possibility of genetic modifiers in SSRIDDs and CdLS pathogenesis and the success of identifying candidate genetic modifiers in Drosophila, Drosophila is an excellent choice for modeling SSRIDDs and CdLS.

## Research Purpose

Overall, the role of genetic background in human disease pathogenesis remains understudied.Genotype-phenotype studies rely on accurate phenotyping and consideration of phenotypic variation across individuals. In Chapter 2, I describe a novel, inexpensive, high-throughput method for quantifying time to sedation in individual Drosophila, a key phenotype for modeling ethanol-related disorders in Drosophila, including FASD. I then take advantage of this new assay for other projects involving ethanol response, now able to consider time to sedation of individual flies and quantify within-line variation.

Until the development of single-cell technologies, investigation of changes in individual cell types within heterogeneous tissue, such as the brain, was challenging. We do not know the degree to which bulk RNA sequencing masks changes in specific cell types.

With single-cell RNA sequencing, we can observe transcriptomic responses to external chemicals, such as ethanol, at a much higher, never-before-seen resolution. In Chapter 3, we take advantage of single-cell technologies and examine the impact of developmental ethanol exposure on the transcriptome of adult Drosophila brain tissue. We show sexually dimorphic responses in subsets of neural and glial cells that were associated with lipid transport, glutamate, GABA, and glutathione metabolism, and vision. Here, I also expand the phenotype associated with developmental ethanol exposure in *Canton S-B* Drosophila, quantifying decreased viability and individual changes to sleep, activity, and ethanol-sedation.

Although previous work reports the impact of genetic background on developmental ethanol exposure, background-dependent differences at the level of the transcriptome as a result of developmental ethanol exposure and how these transcriptomic changes may correlate with behavioral changes remains unknown. Additionally, although changes in sleep and activity as a result of developmental ethanol exposure were reported in *Canton S-B* Drosophila, the extent to which these changes varied with genetic background was unresolved. In Chapter 4, I demonstrate background-dependent effects of developmental ethanol exposure on sleep and activity phenotypes as a part of a larger study on background-dependent changes in gene expression in whole-fly tissue after developmental ethanol exposure as a model for FASD. This study found extensive sexual dimorphism in gene expression and identified a female-specific coordinated network of small nucleolar RNAs (snoRNAs) in response to developmental ethanol exposure.

The role of snoRNAs in ethanol response identified in Chapter 4 led to broader questions about the roles of noncoding RNAs in stress response in Drosophila; compared to protein-coding genes, only a few noncoding RNAs have been comprehensively characterized in Drosophila. The study outlined in Chapter 4 also identified a previously unstudied lncRNA in *D. melanogaster* associated with response to developmental ethanol exposure and host to snoRNAs, *U snoRNA host gene 4* (*Uhg4*). Chapter 5 details the first characterization of the lncRNA and snoRNA host gene *Uhg4,* where I performed behavioral and transcriptomic analysis across in-house generated *Uhg4* deletion mutants and found *Uhg4* to be critical for multiple fitness and stress-response traits, including oocyte development and ethanol sedation time.

Genetic background does not only impact common disease traits, such as FASD, but also rare genetic disorders with varied phenotypic presentations. Varied phenotypes suggest the presence of genetic modifiers naturally segregating in the human population, but such modifiers for rare chromatin modification disorders have not yet been identified. In some diseases with a wide range of phenotypes, disorder classification is shifting towards a spectrum-based approach based on clinical findings. SSRIDDs and CdLS are two neurodevelopmental disorders of chromatin modification with a wide range of within-disorder clinical findings that are shifting towards a spectrum-based classification approach. This shift is purely based on clinical observations; no studies have examined whether molecular evidence for a spectrum-based approach to SSRIDDS and/or CdLS exists. We seek to use Drosophila as a model for SSRIDDs and CdLS, to investigate the relationship between different subtypes of SSRIDDs and CdLS and hypothesize an explanation for the phenotypic variation across these disorders. In Chapter 6, I use the Drosophila system to establish the first RNAi-based co-isogenic Drosophila models for many subtypes of SSRIDDs and CdLS. I quantified gene-specific changes in behavior, brain morphology and the transcriptome, supporting a spectrum-based approach to classification of these disorders. I also identified genes co-regulated with the Drosophila ortholog of SSRIDD-associated *SMARCB1*. The human orthologs of these genes are potential candidate genetic modifiers for *SMARCB1*-SSRIDD.

Out of all genes associated with SSRIDDs, *ARID1B* is the most common. *ARID1B* is also associated with cancers and Autism Spectrum Disorder, but genetic modifiers for *ARID1B*, or an explanation for the wide range of phenotypes associated with *ARID1B*, have not yet been identified. In Chapter 7, I show that there are naturally occurring genetic modifiers for a Drosophila model of *ARID1B*-related SSRIDD using a recently expanded DGRP. I also perform a GWA using a recently expanded DGRP and a Drosophila model of *ARID1B*-related SSRIDD limited to wing tissue, as the fly ortholog of *ARID1B* is critical for survival. Although this chapter has identified potential modifiers, validation of these modifiers is still ongoing.

# References

Adams MD, Holt RA, Scherer SE, Hoskins RA, Galle RF, Henderson SN, Sutton GG, Yandell MD, Zhang Q, Chen LX, et al. The genome sequence of *Drosophila melanogaster*. *Science.* 2000. 287(5461):2185-95.

Alba V, Carthew JE, Carthew RW, Mani M. Global constraints within the developmental program of the Drosophila wing. *eLife.* 2021. 10.

Alfert A, Moreno N, Kerl K. The BAF complex in development and disease. *Epigenetics Chromatin.* 2019. 12(1):19.

Alic N, Hoddinott MP, Foley A, Slack C, Piper MDW, Partridge L. Detrimental effects of RNAi: A cautionary note on its use in Drosophila ageing studies. *PLoS ONE.* 2012. 7(9):e45367.

Al-Jawahiri R, Foroutan A, McConkey H, Levy M, Haghshenas S, Rooney K, Turner J, Shears D, Holder M, Lefroy H, et al. *SOX11* variants cause a neurodevelopmental disorder with infrequent ocular malformations and hypogonadotropic hypogonadism and with distinct DNA methylation profile. *Genet Med.* 2022. 24(6):1261-73.

Allen AM, Neville MC, Birtles S, Croset V, Treiber CD, Waddell S, Goodwin SF. A single-cell transcriptomic atlas of the adult Drosophila ventral nerve cord. *eLife.* 2020. 9.

Alomer RM, da Silva EML, Chen J, Piekarz KM, McDonald K, Sansam CG, Sansam CL, Rankin S. *Esco1* and *Esco2* regulate distinct cohesin functions during cell cycle progression. *Proc Nat Acad Sci U S A.* 2017. 114(37):9906-11.

Aref-Eshghi E, Bend EG, Hood RL, Schenkel LC, Carere DA, Chakrabarti R, Nagamani SCS, Cheung SW, Campeau PM, Prasad C, et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin–Siris and Nicolaides–Baraitser syndromes. *Nat Commun.* 2018. 9(1):4885.

Aref-Eshghi E, Kerkhof J, Pedro VP, France GD, Barat-Houari M, Ruiz-Pallares N, Andrau J, Lacombe D, Van-Gils J, Fergelot P, et al. Evaluation of DNA methylation episignatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. *Am J Hum Genet.* 2021. 108(6):1161-3.

Baker BM, Mokashi SS, Shankar V, Hatfield JS, Hannah RC, Mackay TFC, Anholt RRH. The Drosophila brain on cocaine at single-cell resolution. *Genome Res.* 2021. 31(10):1927-37.

Barish S, Barakat TS, Michel BC, Mashtalir N, Phillips JB, Valencia AM, Ugur B, Wegner J, Scott TM, Bostwick B, et al. *BICRA,* a SWI/SNF complex member, is associated with BAF-disorder related phenotypes in humans and model organisms. *Am J Hum Genet.* 2020. 107(6):1096-112.

Beaty TH, Ruczinski I, Murray JC, Marazita ML, Munger RG, Hetmanski JB, Murray T, Redett RJ, Fallin MD, Liang KY, et al. Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genet Epidemiol.* 2011. 35(6):469-78.

Belhorma K, Darwish N, Benn-Hirsch E, Duenas A, Gates H, Sanghera N, Wu J, French RL. Developmental ethanol exposure causes central nervous system dysfunction and may slow the aging process in a Drosophila model of fetal alcohol spectrum disorder. *Alcohol.* 2021. 94:65-73.

Bell AJ, McBride SMJ, Dockendorff TC. Flies as the ointment: Drosophila modeling to enhance drug discovery. *Fly (Austin).* 2009. 3(1):39-49.

Bellen HJ, Tong C, Tsuda H. 100 years of Drosophila research and its impact on vertebrate neuroscience: A history lesson for the future. *Nat Rev Neurosci.* 2010. 11(7):514-22.

Bellen HJ, Wangler MF, Yamamoto S. The fruit fly at the interface of diagnosis and pathogenic mechanisms of rare and common human diseases. *Hum Mol Genet.* 2019. 28(R2):R207-14.

Berger KH, Heberlein U, Moore MS. Rapid and chronic: Two distinct forms of ethanol tolerance in Drosophila. *Alcohol Clin Exp Res.* 2004. 28(10):1469-80.

Bis-Brewer DM, Fazal S, Züchner S. Genetic modifiers and non-Mendelian aspects of CMT. *Brain Res.* 2020. 1726:146459.

Bjork JM, Gilman JM. The effects of acute alcohol administration on the human brain: Insights from neuroimaging. *Neuropharmacology.* 2014. 84:101-10.

Bögershausen N, Wollnik B. Mutational landscapes and phenotypic spectrum of SWI/SNF-related intellectual disability disorders. *Front Mol Neurosci.* 2018. 11:252.

Boltsis I, Grosveld F, Giraud G, Kolovos P. Chromatin conformation in development and disease. *Front Cell Dev Biol.* 2021. 9:723859.

Bolus H, Crocker K, Boekhoff-Falk G, Chtarbanova S. Modeling neurodegenerative disorders in *Drosophila melanogaster*. *Int J Mol Sci.* 2020. 21(9):3055.

Bonilla M, McPherson M, Coreas J, Boulos M, Chavol P, Alrabadi RI, Loza-Coll M. Repeated ethanol intoxications of *Drosophila melanogaster* adults increases the resistance to ethanol of their progeny. *Alcohol Clin Exp Res.* 2021. 45(7):1370-82.

Bonthius DJ, Winters Z, Karacay B, Bousquet SL, Bonthius DJ. Importance of genetics in fetal alcohol effects: Null mutation of the *nNOS* gene worsens alcohol-induced cerebellar neuronal losses and behavioral deficits. *Neurotoxicol.* 2015. 46:60-72.

Boyles AL, DeRoo LA, Lie RT, Taylor JA, Jugessur A, Murray JC, Wilcox AJ. Maternal alcohol consumption, alcohol metabolism genes, and the risk of oral clefts: A population-based case-control study in Norway, 1996–2001. *Am J Epidemiol.* 2010. 172(8):924-31.

Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* 1993. 118(2):401-15.

Brenner E, Tiwari GR, Kapoor M, Liu Y, Brock A, Mayfield RD. Single cell transcriptome profiling of the human alcohol-dependent brain. *Hum Mol Genet.* 2020. 29(7):1144-53.

Brocardo PS, Gil-Mohapel J, Christie BR. The role of oxidative stress in fetal alcohol spectrum disorders. *Brain Res Rev.* 2011. 67(1):209-25.

Campbell SS, Tobler I. Animal sleep: A review of sleep duration across phylogeny. *Neurosci Biobehav Rev.* 1984. 8(3):269-300.

Cavallin M, Rujano MA, Bednarek N, Medina-Cano D, Bernabe Gelot A, Drunat S, Maillard C, Garfa-Traore M, Bole C, Nitschké P, et al. *WDR81* mutations cause extreme microcephaly and impair mitotic progression in human fibroblasts and Drosophila neural stem cells. *Brain.* 2017. 140(10):2597-609.

Cenik BK, Shilatifard A. COMPASS and SWI/SNF complexes in development and disease. *Nat Rev Genet.* 2021. 22(1):38-58.

Centore RC, Sandoval GJ, Soares LMM, Kadoch C, Chan HM. Mammalian SWI/SNF chromatin remodeling complexes: Emerging mechanisms and therapeutic strategies. *Trends Genet.* 2020. 36(12):936-50.

Chatterjee D, Costa CAM, Wang X, Jevitt A, Huang Y, Deng W. Single-cell transcriptomics identifies Keap1-Nrf2 regulated collective invasion in a Drosophila tumor model. *eLife.* 2022. 11.

Chien S, Reiter LT, Bier E, Gribskov M. Homophila: Human disease gene cognates in Drosophila. *Nucleic Acids Res.* 2002. 30(1):149-51.

Chubak MC, Nixon KCJ, Stone MH, Raun N, Rice SL, Sarikahya M, Jones SG, Lyons TA, Jakub TE, Mainland RLM, et al. Individual components of the SWI/SNF chromatin remodelling complex have distinct roles in memory neurons of the Drosophila mushroom body. *Dis Model Mech.* 2019. 12(3):dmm037325.

Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet.* 2014. 15(1):34-48.

Coffin GS, Siris E. Mental retardation with absent fifth fingernail and terminal phalanx. *Am J Dis Child.* 1970. 119(5):433-9.

Cohan FM, Graf JD. Latitudinal cline in *Drosophila melanogaster* for knockdown resistance to ethanol fumes and for rates of response to selection for further resistance. *Evolution.* 1985. 39(2):278-93.

Coll-Tane M, Krebbers A, Castells-Nobau A, Zweier C, Schenck A. Intellectual disability and autism spectrum disorders 'on the fly': Insights from Drosophila. *Dis Model Mech.* 2019. 12(5):dmm039180. doi: 10.1242/dmm.039180.

Cukier HN, Perez AM, Collins AL, Zhou Z, Zoghbi HY, Botas J. Genetic modifiers of MeCP2 function in Drosophila. *PLoS Genet.* 2008. 4(9):e1000179.

Dalton HM, Viswanatha R, Brathwaite J, Roderick, Zuno JS, Berman AR, Rushforth R, Mohr SE, Perrimon N, Chow CY. A genome-wide CRISPR screen identifies *DPM1* as a modifier of DPAGT1 deficiency and ER stress. *PLoS Genet.* 2022. 18(9):e1010430.

Das UG, Cronk CE, Martier SS, Simpson PM, McCarver DG. Alcohol dehydrogenase 2*3 affects alterations in offspring facial morphology associated with maternal ethanol intake in pregnancy. *Alcohol Clin Exp Res.* 2004. 28(10):1598-606.

Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft Ł, Aibar S, Makhzami S, Christiaens V, Bravo González-Blas C, et al. A single-cell transcriptome atlas of the aging Drosophila brain. *Cell.* 2018. 174(4):982,998.e20.

Deardorff MA, Noon SE, Krantz ID. 2020. Cornelia de Lange syndrome. In: Gene reviews. Adam MP, Mirzaa GM, Pagon RA,et al., editors. Seattle: University of Washington.

Deardorff MA, Bando M, Nakato R, Watrin E, Itoh T, Minamino M, Saitoh K, Komata M, Katou Y, Clark D, et al. *HDAC8* mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature.* 2012. 489:313-7.

Devineni AV, Heberlein U. The evolution of *Drosophila melanogaster* as a model for alcohol research. *Annu Rev Neurosci.* 2013. 36(1):121-38.

Dexter JS. The analysis of a case of continuous variation in Drosophila by a study of its linkage relations. *Am Nat.* 1914. 48(576):712-58.

Di Nardo M, Pallotta MM, Musio A. The multifaceted roles of cohesin in cancer. *J Exp Clin Cancer Res.* 2022. 41(1):96.

Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature.* 2007. 448(7150):151-6.

Dodge NC, Jacobson JL, Jacobson SW. Protective effects of the alcohol dehydrogenase-ADH1B*3 allele on attention and behavior problems in adolescents exposed to alcohol during pregnancy. *Neurotoxicol and Teratol.* 2014. 41:43-50.

Dorsett D, Krantz ID. On the molecular etiology of Cornelia de Lange syndrome. *Ann N Y Acad Sci.* 2009. 1151:22-37.

Dorsett D. The *Drosophila melanogaster* model for Cornelia de Lange syndrome: Implications for etiology and therapeutics. *Am J Med Genet.* 2016. 172(2):129-37.

Dubowy C, Sehgal A. Circadian rhythms and sleep in *Drosophila melanogaster*. *Genetics.* 2017. 205(4):1373-97.

Eberhart JK, Parnell SE. The genetics of fetal alcohol spectrum disorders. *Alcohol Clin Exp Res.* 2016. 40(6):1154-65.

Elvig SK, McGinn MA, Smith C, Arends MA, Koob GF, Vendruscolo LF. Tolerance to alcohol: A critical yet understudied factor in alcohol addiction. *Pharmacol Biochem Behav.* 2021. 204:173155.

Engel GL, Marella S, Kaun KR, Wu J, Adhikari P, Kong EC, Wolf FW. Sir2/Sirt1 links acute inebriation to presynaptic changes and the development of alcohol tolerance, preference, and reward. *J Neurosci.* 2016. 36(19):5241-51.

Engel GL, Taber K, Vinton E, Crocker AJ. Studying alcohol use disorder using *Drosophila melanogaster* in the era of ‘Big data’. *Behav Brain Funct.* 2019. 15(1):7.

Evangelou A, Ignatiou A, Antoniou C, Kalanidou S, Chatzimatthaiou S, Shianiou G, Ellina S, Athanasiou R, Panagi M, Apidianakis Y, et al. Unpredictable effects of the genetic background of transgenic lines in physiological quantitative traits. *G3 (Bethesda).* 2019. 9(11):3877-90.

Finelli R, Mottola F, Agarwal A. Impact of alcohol consumption on male fertility potential: A narrative review. *Int J Environ Res Public Health.* 2021. 19(1):328.

Fochler S, Morozova TV, Davis MR, Gearhart AW, Huang W, Mackay TFC, Anholt RRH. Genetics of alcohol consumption in *Drosophila melanogaster*. *Genes Brain Behav.* 2017. 16(7):675-85.

Ghezzi A, Krishnan HR, Atkinson NS. Susceptibility to ethanol withdrawal seizures is produced by BK channel gene expression. *Addict Biol.* 2014. 19(3):332-7.

Global Genes. Rare Disease Facts [Internet] [cited 2023 Mar 22]. Available from: https://globalgenes.org/rare-disease-facts/ .

Goh ES, Li C, Horsburgh S, Kasai Y, Kolomietz E, Morel CF. The Roberts syndrome/SC phocomelia spectrum—A case report of an adult with review of the literature. *Am J Med Genet. A.* 2010. 152A(2):472-8.

Gosdin LK, Deputy NP, Kim SY, Dang EP, Denny CH. Alcohol consumption and binge drinking during pregnancy among adults aged 18–49 years — United States, 2018–2020. *MMWR. Morbidity and Mortality Weekly Report.* 2022. 71(1):10-3.

Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, dos Santos G, Goodman JL, Goutte-Gattat D, Jenkins VK, Kaufman T, et al. FlyBase: A guided tour of highlighted features. *Genetics.* 2022. 220(4).

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics.* 2013. 194(4):1029-35.

Green EW, Fedele G, Giorgini F, Kyriacou CP. A Drosophila RNAi collection is subject to dominant phenotypic effects. *Nat Methods.* 2014. 11(3):222-3.

Guevara A, Gates H, Urbina B, French R. Developmental ethanol exposure causes reduced feeding and reveals a critical role for Neuropeptide F in survival. *Front Physiol.* 2018. 9:237.

Guzman DM, Chakka K, Shi T, Marron A, Fiorito AE, Rahman NS, Ro S, Sucich DG, Pierce JT. Transgenerational effects of alcohol on behavioral sensitivity to alcohol in C*aenorhabditis elegans*. *PloS ONE.* 2022. 17(10):e0271849.

Harbison ST, McCoy LJ, Mackay TFC. Genome-wide association study of sleep in *Drosophila melanogaster*. *BMC Genomics.* 2013. 14(1):281.

Harbison ST, Serrano Negron YL, Hansen NF, Lobell AS. Selection for long and short sleep duration in *Drosophila melanogaster* reveals the complex genetic network underlying natural variation in sleep. *PLoS Genet.* 2017. 13(12):e1007098.

Harbison ST, Kumar S, Huang W, McCoy LJ, Smith KR, Mackay TFC. Genome-wide association study of circadian behavior in *Drosophila melanogaster*. *Behav Genet.* 2019. 49(1):60-82.

Hasin DS, Grant BF. The national epidemiologic survey on alcohol and related conditions (NESARC) waves 1 and 2: Review and summary of findings. *Soc Psychiatry Psychiatr Epidemiol.* 2015. 50(11):1609-40.

He S, Wu Z, Tian Y, Yu Z, Yu J, Wang X, Li J, Liu B, Xu Y. Structure of nucleosome-bound human BAF complex. *Science.* 2020. 367(6480):875-81.

He X, Zhou S, St Armour GE, Mackay TF, Anholt RR. Epistatic partners of neurogenic genes modulate Drosophila olfactory behavior. *Genes Brain Behav.* 2016. 15(2):280-90.

Heigwer F, Port F, Boutros M. RNA interference (RNAi) screening in Drosophila. *Genetics.* 2018. 208(3):853-74.

Hong M, Krauss RS. Cdon mutation and fetal ethanol exposure synergize to produce midline signaling defects and holoprosencephaly spectrum disorders in mice. *PLoS Genet.* 2012. 8(10):e1002999.

Hong M, Krauss RS. Rescue of holoprosencephaly in fetal alcohol-exposed Cdon mutant mice by reduced gene dosage of Ptch1. *PLoS ONE.* 2013. 8(11):e79269.

Hope KA, Berman AR, Peterson RT, Chow CY. An *in vivo* drug repurposing screen and transcriptional analyses reveals the serotonin pathway and GSK3 as major therapeutic targets for NGLY1 deficiency. *PLoS Genet.* 2022. 18(6):e1010228.

Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, et al. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.* 2015. 25(3):445-58.

Huang W, Carbone MA, Lyman RF, Anholt RRH, Mackay TFC. Genotype by environment interaction for gene expression in *Drosophila melanogaster*. *Nat Commun.* 2020. 11(1):5451.

Huber R, Hill SL, Holladay C, Biesiadecki M, Tononi G, Cirelli C. Sleep homeostasis in *Drosophila melanogaster*. *Sleep.* 2004. 27(4):628-39.

Inkelis SM, Thomas JD. Sleep in infants and children with prenatal alcohol exposure. *Alcohol Clin Exp Res.* 2018. 42(8):1390-405.

Isaac RE, Li C, Leedale AE, Shirras AD. Drosophila male sex peptide inhibits siesta sleep and promotes locomotor activity in the post-mated female. *Proc R Soc B.* 2010. 277(1678):65-70.

Ja WW, Carvalho GB, Mak EM, de la Rosa NN, Fang AY, Liong JC, Brummel T, Benzer S. Prandiology of Drosophila and the CAFE assay. *Proc Nat Acad Sci U S A.* 2007. 104(20):8253-6.

Jacobson SW, Carr LG, Croxford J, Sokol RJ, Li T, Jacobson JL. Protective effects of the alcohol dehydrogenase- ADH1B allele in children exposed to alcohol during pregnancy. *J Pediatr.* 2006. 148(1):30-7.

Janssens J, Aibar S, Taskiran II, Ismail JN, Gomez AE, Aughey G, Spanier KI, De Rop FV, González-Blas CB, Dionne M, et al. Decoding gene regulation in the fly brain. *Nature.* 2022. 601(7894):630-6.

Jenett A, Rubin G, Ngo TB, Shepherd D, Murphy C, Dionne H, Pfeiffer B, Cavallaro A, Hall D, Jeter J, et al. A GAL4-driver line resource for Drosophila neurobiology. *Cell Rep.* 2012. 2(4):991-1001.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA—Guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012. 337(6096):816-21.

Joiner WJ, Crocker A, White BH, Sehgal A. Sleep in Drosophila is regulated by adult mushroom bodies. *Nature.* 2006. 441(7094):757-60.

Kadoch C, Crabtree GR. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv.* 2015. 1(5):e1500447.

Kadoch C, Williams RT, Calarco JP, Miller EL, Weber CM, Braun SMG, Pulice JL, Chory EJ, Crabtree GR. Dynamics of BAF-polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat Genet.* 2017. 49(2):213-22.

Kang H, Zhao J, Jiang X, Li G, Huang W, Cheng H, Duan R. Drosophila *Netrin-B* controls mushroom body axon extension and regulates courtship-associated learning and memory of a Drosophila Fragile X syndrome model. *Mol Brain.* 2019. 12(1):52.

Karacay B, Bonthius NE, Plume J, Bonthius DJ. Genetic absence of nNOS worsens fetal alcohol effects in mice. I: Behavioral deficits. *Alcohol Clin Exp Res.* 2015. 39(2):212-20.

Karacay B, Mahoney J, Plume J, Bonthius DJ. Genetic absence of nNOS worsens fetal alcohol effects in mice. II: Microencephaly and neuronal losses. *Alcohol Clin Exp Res.* 2015. 39(2):221-31.

Kim S, Kim J, Park S, Park J, Lee S. Drosophila Graf regulates mushroom body *β*-axon extension and olfactory long-term memory. *Mol Brain.* 2021. 14(1):73.

Kleefstra T, Kramer J, Neveling K, Willemsen M, Koemans T, Vissers LLM, Wissink-Lindhout W, Fenckova M, van den Akker WR, Kasri N, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am J Hum Genet.* 2012. 91(1):73-82.

Kleiber ML, Diehl EJ, Laufer BI, Mantha K, Chokroborty-Hoque A, Alberry B, Singh SM. Long-term genomic and epigenomic dysregulation as a consequence of prenatal alcohol exposure: A model for fetal alcohol spectrum disorders. *Front Genet.* 2014. 5:161.

Kline AD, Moss JF, Selicorni A, Bisgaard A, Deardorff MA, Gillett PM, Ishman SL, Kerr LM, Levin AV, Mulder PA, et al. Diagnosis and management of Cornelia de Lange syndrome: First international consensus statement. *Nat Rev Genet.* 2018. 19(10):649-66.

Koh K, Evans JM, Hendricks JC, Sehgal A. A Drosophila model for age-associated changes in sleep:Wake cycles. *Proc Nat Acad Sci U S A.* 2006. 103(37):13843-7.

Kramer JM, Kochinke K, Oortveld MAW, Marks H, Kramer D, Jong EKd, Asztalos Z, Westwood JT, Stunnenberg HG, Sokolowski MB, et al. Epigenetic regulation of learning and memory by Drosophila EHMT/G9a. *PLoS Biol.* 2011. 9(1):e1000569.

Lange S, Probst C, Gmel G, Rehm J, Burd L, Popova S. Global prevalence of fetal alcohol spectrum disorder among children and youth: A systematic review and meta-analysis. *JAMA Pediatr.* 2017. 171(10):948-56.

Lavoy S, Chittoor-Vinod V, Chow CY, Martin I. Genetic modifiers of neurodegeneration in a Drosophila model of Parkinson's disease. *Genetics.* 2018. 209(4):1345-56.

Li F, Lindsey JW, Marin EC, Otto N, Dreher M, Dempsey G, Stark I, Bates AS, Pleijzier MW, Schlegel P, et al. The connectome of the adult Drosophila mushroom body provides insights into function. *eLife.* 2020. 9.

Li Y, Yang H, Zdanowicz M, Sicklick JK, Qi Y, Camp TJ, Diehl AM. Fetal alcohol exposure impairs hedgehog cholesterol modification and signaling. *Lab Invest.* 2007. 87(3):231-40.

Li Y, Seto E. HDACs and HDAC inhibitors in cancer development and therapy. *Cold Spring Harb Perspect Med.* 2016. 6(10):a026831.

Link N, Bellen HJ. Using Drosophila to drive the diagnosis and understand the mechanisms of rare human diseases. *Development.* 2020. 147(21).

Liu S, Liu Q, Tabuchi M, Wu MN. Sleep drive is encoded by neural plastic changes in a dedicated circuit. *Cell.* 2016. 165(6):1347-60.

Logan-Garbisch T, Bortolazzo A, Luu P, Ford A, Do D, Khodabakhshi P, French RL. Developmental ethanol exposure leads to dysregulation of lipid metabolism and oxidative stress in Drosophila. *G3 (Bethesda).* 2015. 5(1):49-59.

Luhur A, Klueg KM, Zelhof AC. Generating and working with Drosophila cell cultures: Current challenges and opportunities. *Wiley Interdiscip Rev Dev Biol.* 2019. 8(3):e339.

Lupton C, Burd L, Harwood R. Cost of fetal alcohol spectrum disorders. *Am J Med Genet C Semin Med Genet.* 2004. 127C(1):42-50.

Ly S, Pack AI, Naidoo N. The neurobiological basis of sleep: Insights from Drosophila. *Neurosci Biobehav Rev.* 2018. 87:67-86.

Ma Y, Creanga A, Lum L, Beachy PA. Prevalence of off-target effects in Drosophila RNA interference screens. *Nature.* 2006. 443(7109):359-63.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. The *Drosophila melanogaster* genetic reference panel. *Nature.* 2012. 482(7384):173-8.

Mackay TFC. Epistasis and quantitative traits: Using model organisms to study gene-gene interactions. *Nat Rev Genet.* 2014. 15(1):22-33.

Mackay TFC, Huang W. Charting the genotype–phenotype map: Lessons from the *Drosophila melanogaster* genetic reference panel. *Wiley Interdiscip Rev Dev Biol.* 2018. 7(1):e289.

Marine J, Dawson S, Dawson MA. Non-genetic mechanisms of therapeutic resistance in cancer. *Nat Rev Cancer*2020. 20(12):743-56.

Martin F, Sánchez-Hernández S, Gutiérrez-Guerrero A, Pinedo-Gomez J, Benabdellah K. Biased and unbiased methods for the detection of off-target cleavage by CRISPR/Cas9: An overview. *Int J Mol Sci.* 2016. 17(9):1507.

Marwaha S, Knowles JW, Ashley EA. A guide for the diagnosis of rare and undiagnosed disease: Beyond the exome. *Genome Med.* 2022. 14(1):23.

May PA, Baete A, Russo J, Elliott AJ, Blankenship J, Kalberg WO, Buckley D, Brooks M, Hasken J, Abdul-Rahman O, et al. Prevalence and characteristics of fetal alcohol spectrum disorders. *Pediatrics.* 2014. 134(5):855-66.

McCarthy N, Wetherill L, Lovely CB, Swartz ME, Foroud TM, Eberhart JK. Pdgfra protects against ethanol-induced craniofacial defects in a zebrafish model of FASD. *Development.* 2013. 140(15):3254-65.

McCarver DG, Thomasson HR, Martier SS, Sokol RJ, Li T. Alcohol dehydrogenase-2*3 allele protects against alcohol-related birth defects among African Americans. *J Pharmacol Exp Ther.* 1997. 283(3):1095-101.

McClure KD, French RL, Heberlein U. A Drosophila model for fetal alcohol syndrome disorders: Role for the insulin pathway. *Dis Model Mech.* 2011. 4(3):335-46.

Michel BC, D'Avino AR, Cassel SH, Mashtalir N, McKenzie ZM, McBride MJ, Valencia AM, Zhou Q, Bocker M, Soares LMM, et al. A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol.* 2018. 20(12):1410-20.

Mirzoyan Z, Sollazzo M, Allocca M, Valenza AM, Grifoni D, Bellosta P. *Drosophila melanogaster*: A model organism to study cancer. *Front Genet.* 2019. 10:51.

Misulovin Z, Schwartz YB, Li X, Kahn TG, Gause M, MacArthur S, Fay JC, Eisen MB, Pirrotta V, Biggin MD, et al. Association of cohesin and *Nipped-B* with transcriptionally active regions of the *Drosophila melanogaster* genome. *Chromosoma.* 2008. 117(1):89-102.

Modi MN, Shuai Y, Turner GC. The Drosophila mushroom body: From architecture to algorithm in a learning circuit. *Annu Rev Neurosci.* 2020. 43(1):465-84.

Moore EM, Riley EP. What happens when children with fetal alcohol spectrum disorders become adults? *Curr Dev Disord Rep.* 2015. 2(3):219-27.

Morgan TH. Sex limited inheritance in Drosophila. *Science.* 1910. 32(812):120-2.

Morozova TV, Anholt RRH, Mackay TFC. Transcriptional response to alcohol exposure in *Drosophila melanogaster*. *Genome Biol.* 2006. 7(10):R95.

Morozova TV, Huang W, Pray VA, Whitham T, Anholt RRH, Mackay TFC. Polymorphisms in early neurodevelopmental genes affect natural variation in alcohol sensitivity in adult Drosophila. *BMC Genomics.* 2015. 16(1):865.

Morozova TV, Hussain Y, McCoy LJ, Zhirnov EV, Davis MR, Pray VA, Lyman RA, Duncan LH, McMillen A, Jones A, et al. A *Cyclin E* centered genetic network contributes to alcohol-induced variation in Drosophila development. *G3 (Bethesda).* 2018. 8(8):2643-53.

Nasmyth K, Haering CH. Cohesin: Its roles and mechanisms. *Annu Rev Genet.* 2009. 43(1):525-58.

National Center for Chronic Disease Prevention and Health Promotion. Excessive Alcohol Use [Internet]; c2022 [cited 2023 Mar 8.]. Available from: https://www.cdc.gov/chronicdisease/resources/publications/factsheets/alcohol.htm .

National Center on Birth Defects and Developmental Disabilities. Fetal Alcohol Spectrum Disorders (FASDs) [Internet]; c2022 [cited 2023 Mar 7,]. Available from: https://www.cdc.gov/ncbddd/fasd/facts.html .

National Human Genome Res Institute. Rare Diseases FAQ [Internet]; c2020 [cited 2023 Mar 22,]. Available from: https://www.genome.gov/FAQ/Rare-Diseases .

National Institute on Alcohol Abuse and Alcoholism. Understanding Alcohol Use Disorder [Internet]; c2021 [cited 2023 Mar 8,]. Available from: https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/understanding-alcohol-use-disorder .

Nicolaides P, Baraitser M. An unusual syndrome with mental retardation and sparse hair. *Clin Dysmorphol.* 1993. 2(3):232-6.

Nixon KCJ, Rousseau J, Stone MH, Sarikahya M, Ehresmann S, Mizuno S, Matsumoto N, Miyake N, Baralle D, McKee S, et al. A syndromic neurodevelopmental disorder caused by mutations in SMARCD1, a core SWI/SNF subunit needed for context-dependent neuronal gene regulation in flies. *Am J Hum Genet.* 2019. 104(4):596-610.

Özsoy ED, Yılmaz M, Patlar B, Emecen G, Durmaz E, Magwire MM, Zhou S, Huang W, Anholt RRH, Mackay TFC. Epistasis for head morphology in *Drosophila melanogaster*. *G3 (Bethesda).* 2021. 11(10).

Palu RAS, Ong E, Stevens K, Chung S, Owings KG, Goodman AG, Chow CY. Natural genetic variation screen in Drosophila identifies Wnt signaling, mitochondrial metabolism, and redox homeostasis genes as modifiers of apoptosis. *G3 (Bethesda).* 2019. 9(12):3995-4005.

Parenti I, Kaiser FJ. Cornelia de Lange syndrome as paradigm of chromatinopathies. *Front Neurosci.* 2021. 15:774950.

Pauli A, Althoff F, Oliveira RA, Heidmann S, Schuldiner O, Lehner CF, Dickson BJ, Nasmyth K. Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. *Dev Cell.* 2008. 14(2):239-51.

Pauli A, van Bemmel JG, Oliveira RA, Itoh T, Shirahige K, van Steensel B, Nasmyth K. A direct role for cohesin in gene regulation and ecdysone response in Drosophila salivary glands. *Curr Biol.* 2010. Oct 26;20(20):1787-98.

Perkins LA, Holderbaum L, Tao R, Hu Y, Sopko R, McCall K, Yang-Zhou D, Flockhart I, Binari R, Shim H, et al. The transgenic RNAi project at Harvard medical school: Resources and validation. *Genetics.* 2015. 201(3):843-52.

PBC Perlara, Mayo Clinic and University of Utah. *MAN1B1*-CDG Cure Odyssey [Internet]; c2023 [cited 2023 Mar 21]. Available from: https://perlara.substack.com/p/man1b1-cdg-cure-odyssey .

Petrelli B, Weinberg J, Hicks GG. Effects of prenatal alcohol exposure (PAE): Insights into FASD using mouse models of PAE. *Biochem Cell Biol.* 2018. 96(2):131-47.

Petruccelli E, Kaun KR. Insights from intoxicated Drosophila. *Alcohol.* 2019. 74:21-7.

Pimentel D, Donlea JM, Talbot CB, Song SM, Thurston AJF, Miesenböck G. Operation of a homeostatic sleep switch. *Nature.* 2016. 536(7616):333-7.

Pitchers W, Nye J, Márquez EJ, Kowalski A, Dworkin I, Houle D. A multivariate genome-wide association study of wing shape in *Drosophila melanogaster*. *Genetics.* 2019. 211(4):1429-47.

Pitman J, McGill J, Keegan K, Allada R. A dynamic role for the mushroom bodies in promoting sleep in Drosophila. *Nature.* 2006. 441(7094):753-6.

Quinn WG, Harris WA, Benzer S. Conditioned behavior in *Drosophila melanogaster*. *Proc Nat Acad Sci U S A.* 1974. 71(3):708-12.

Rahit KMTH, Tarailo-Graovac M. Genetic modifiers and rare mendelian disease. *Genes.* 2020. 11(3):239.

Reiter LT, Potocki L, Chien S, Gribskov M, Bier E. A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res.* 2001. 11(6):1114-25.

Romitti PA, Lidral AC, Munger RG, Daack-Hirsch S, Burns TL, Murray JC. Candidate genes for nonsyndromic cleft lip and palate and maternal cigarette smoking and alcohol consumption: Evaluation of genotype-environment interactions from a population-based case-control study of orofacial clefts. *Teratology.* 1999. 59(1):39-50.

Santen GWE, Aten E, Vulto-van Silfhout AT, Pottinger C, van Bon BWM, van Minderhout, Ivonne J. H. M., Snowdowne R, van der Lans, Christian A. C., Boogaard M, Linssen MML, et al. Coffin–Siris syndrome and the BAF complex: Genotype–Phenotype study in 63 patients. *Hum Mut.* 2013. 34(11):1519-28.

Schrier Vergano S, Santen G, Wieczorek D, Wollnik B, Matsumo N, Deardorff MA. 2013. Coffin-Siris syndrome. In: *GeneReviews* (R). Adam MP, Everman DB, Mirzaa GM,et al., editors. Seattle, WA: University of Washington, Seattle.

Schrier SA, Sherer I, Deardorff MA, Clark D, Audette L, Gillis L, Kline AD, Ernst L, Loomes K, Krantz ID, et al. Causes of death and autopsy findings in a large study cohort of individuals with Cornelia de Lange syndrome and review of the literature. *Am J Med Genet A.* 2011. 155A(12):3007-24.

Schuldiner O, Berdnik D, Levy JM, Wu JS, Luginbuhl D, Gontang AC, Luo L. *piggyBac*-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Dev Cell.* 2008. 14(2):227-38.

Scriver CR, Waters PJ. Monogenic traits are not simple: Lessons from Phenylketonuria. *Trends Genet.* 1999. 15(7):267-72.

Shimaji K, Tanaka R, Maeda T, Ozaki M, Yoshida H, Ohkawa Y, Sato T, Suyama M, Yamaguchi M. Histone methyltransferase G9a is a key regulator of the starvation-induced behaviors in *Drosophila melanogaster*. *Sci Rep.* 2017. 7(1):14763-13.

Sokpor G, Xie Y, Rosenbusch J, Tuoc T. Chromatin remodeling BAF (SWI/SNF) complexes in neural development and disorders. *Front Mol Neurosci.* 2017. 10:243.

Straub J, Konrad EDH, Grüner J, Toutain A, Bok LA, Cho MT, Crawford HP, Dubbs H, Douglas G, Jobling R, et al. Missense variants in *RHOBTB2* cause a developmental and epileptic encephalopathy in humans, and altered levels cause neurological defects in Drosophila. *Am J Hum Genet.* 2018. 102(1):44-57.

Streissguth AP, Dehaene P. Fetal alcohol syndrome in twins of alcoholic mothers: Concordance of diagnosis and IQ. *Am J Med Genet.* 1993. 47(6):857-61.

Substance Abuse and Mental Health Services Administration. 2019 National survey on drug use and health. Center for Behavioral Health Statistics and Quality.

Swarup S, Harbison ST, HAahn LE, Morozova TV, Yamamoto A, Mackay TFC, Anholt RRH. Extensive epistasis for olfactory behaviour, sleep and waking activity in *Drosophila melanogaster*. *Genet Res (Camb).* 2012. 94(1):9-20.

Talsness DM, Owings KG, Coelho E, Mercenne G, Pleinis JM, Partha R, Hope KA, Zuberi AR, Clark NL, Lutz CM, et al. A Drosophila screen identifies NKCC1 as a modifier of NGLY1 deficiency. *eLife.* 2020. 9.

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 2019. 20(8):467-84.

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, et al. COSMIC: The catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 2019. 47(D1):D941-7.

Tawa EA, Hall SD, Lohoff FW. Overview of the genetics of alcohol use disorder. *Alcohol Alcohol.* 2016. 51(5):507-14.

Ugur B, Chen K, Bellen HJ. Drosophila tools and assays for the study of human diseases. *Dis Model Mech.* 2016. 9(3):235-44.

Viljoen DL, Carr LG, Foroud TM, Brooke L, Ramsay M, Li TK. Alcohol dehydrogenase-2*2 allele is associated with decreased prevalence of fetal alcohol syndrome in the mixed-ancestry population of the western cape province, south africa. *Alcohol Clin Exp Res.* 2001. 25(12):1719-22.

Vissers JHA, Manning SA, Kulkarni A, Harvey KF. A Drosophila RNAi library modulates Hippo pathway-dependent tissue growth. *Nat Commun.* 2016. 7(1):10368.

Vonesch S, Lamparter D, Mackay T, Bergmann S, Hafen E. Genome-wide analysis reveals novel regulators of growth in *Drosophila melanogaster*. *PLoS Genet.* 2016. 12(1).

Weber KE. An apparatus for measurement of resistance to gas-phase agents. *Drosophila Information Service.* 1988. 67:91-3.

Wentzel P, Eriksson UJ. Genetic influence on dysmorphogenesis in embryos from different rat strains exposed to ethanol in vivo and in vitro. *Alcohol Clin Exp Res.* 2008. 32(5):874-87.

Wieczorek D, Bögershausen N, Beleggia F, Steiner-Haldenstätt S, Pohl E, Li Y, Milz E, Martin M, Thiele H, Altmüller J, et al. A comprehensive molecular study on Coffin–Siris and Nicolaides–Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. *Hum Mol Genet.* 2013. 22(25):5121-35.

Wilson DA, Masiello K, Lewin MP, Hui M, Smiley JF, Saito M. Developmental ethanol exposure-induced sleep fragmentation predicts adult cognitive impairment. *Neuroscience.* 2016. 322:18-27.

Wu D, Cederbaum AI. Alcohol, oxidative stress, and free radical damage. *Alcohol Res Health.* 2003. 27(4):277-84.

Wu KJ, Kumar S, Serrano Negron YL, Harbison ST. Genotype influences day-to-day variability in sleep in *Drosophila melanogaster*. *Sleep.* 2018. 41(2):1.

Wu Y, Gause M, Xu D, Misulovin Z, Schaaf CA, Mosarla RC, Mannino E, Shannon M, Jones E, Shi M, et al. Drosophila *Nipped-B* mutants model Cornelia de Lange syndrome in growth and behavior. *PLoS Genet.* 2015. 11(11):e1005655.

Xu B, Lee K, Zhang L, Gerton J. Stimulation of mTORC1 with L-leucine rescues defects associated with Roberts syndrome. *PLoS Genet.* 2013. 9(10).

Yamamoto S, Jaiswal M, Charng W, Gambin T, Karaca E, Mirzaa G, Wiszniewski W, Sandoval H, Haelterman N, Xiong B, et al. A Drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell.* 2014. 159(1):200-14.

Zhang C, Ojiaku P, Cole GJ. Forebrain and hindbrain development in zebrafish is sensitive to ethanol exposure involving agrin, Fgf, and sonic hedgehog function. *Birth Defects Res A Clin Mol Teratol.* 2013. 97(1):8-27.

Zhang Y, Wang D, Peng M, Tang L, Ouyang J, Xiong F, Guo C, Tang Y, Zhou Y, Liao Q, et al. Single-cell RNA sequencing in cancer research. *J Exp Clin Cancer Res.* 2021. 40(1):81.

Zhao S, Allis CD, Wang GG. The language of chromatin modification in human cancers. *Nat Rev Cancer*2021. 21(7):413-30.

Zhu Z, Wang X. Roles of cohesin in chromosome architecture and gene expression. *Semin Cell Dev Biol.* 2019. 90:187-93.

Zirin J, Bosch J, Viswanatha R, Mohr SE, Perrimon N. State-of-the-art CRISPR for in vivo and cell-based studies in Drosophila. *Trends Genet.* 2022. 38(5):437-53.

Zwarts L, Vanden Broeck L, Cappuyns E, Ayroles JF, Magwire MM, Vulsteke V, Clements J, Mackay TFC, Callaerts P. The genetic basis of natural variation in mushroom body size in *Drosophila melanogaster*. *Nat Commun.* 2015. 6(1):10115.

# CHAPTER TWO

# HIGH-THROUGHPUT METHOD FOR MEASURING ALCOHOL SEDATION TIME OF INDIVIDUAL *DROSOPHILA MELANOGASTER*

Sass, T.N.*, MacPherson, R.A.*, Mackay, T.F.C., Anholt, R.R.H. High-Throughput Method for Measuring Alcohol Sedation Time of Individual *Drosophila melanogaster*. *J. Vis. Exp.* (158), e61108, doi:10.3791/61108 (2020).
*equal contribution

## Author Contribution Statement

My contributions to the work described in this chapter are as follows: I conceived and developed the assay alongside Tatum Sass, who reared the flies and performed data collection. I performed the data analysis and wrote the first draft of the manuscript, with input from all authors.

## Introduction

The National Institute on Alcohol Abuse and Alcoholism reports that in 2015 excessive alcohol consumption, designated as "alcohol use disorder", affected an estimated 16 million people in the United States. Alcohol abuse causes a wide range of adverse physiological effects and is a major cause of death in the U.S. In humans, decreased sensitivity, or a low level of response to alcohol, has a strong genetic component and is associated with a higher risk of developing alcohol use disorders[1,2,3,4]. Genetic risk studies on human populations are challenging because of population admixture, diverse developmental histories and environmental exposures, and reliance on self-reported questionnaires to quantify alcohol-related phenotypes, which are often confounded with other neuropsychiatric conditions.

*Drosophila melanogaster* provides an excellent model to study the genetic underpinnings of alcohol sensitivity[5,6,7,8]. The *Drosophila* model allows strict control over genetic background, and virtually unlimited numbers of individuals of the same genotype can be reared rapidly under well-controlled environmental conditions without regulatory restrictions and at relatively low cost. In addition to publicly available mutations and RNAi lines that target a majority of genes in the genome, the availability of the *Drosophila melanogaster* Genetic Reference Panel (DGRP), a population of 205 inbred wild-derived lines with complete genome sequences, has enabled genome-wide association studies[9,10]. Such studies have identified genetic networks associated with effects on development time and viability upon developmental exposure to ethanol[11,12]. Evolutionary conservation of fundamental biological processes enables translational inferences to be drawn by superimposing human orthologs on their fly counterparts.

Flies exposed to ethanol undergo physiological and behavioral changes that resemble human alcohol intoxication, including loss of postural control[8], sedation, and development of tolerance[13,14,15]. Alcohol induced sedation in *Drosophila* can be quantified using inebriometers. These are 122 cm long vertical glass columns with slanted mesh partitions to which flies can attach[16,17,18]. A group of at least 50 flies (sexes can be analyzed separately) are introduced in the top of the column and exposed to ethanol vapors. Flies that lose postural control fall through the column and are collected at 1 min intervals. The mean elution time serves as a measure of sensitivity to alcohol intoxication. When flies are exposed to alcohol a second time after recovering from the first exposure, they can develop tolerance, as evident from a shift in mean elution time[13,15,19,20]. Whereas inebriometer assays have led to identification of genes, genetic networks, and cellular pathways associated with alcohol sedation sensitivity and development of tolerance[12,13,14,21], the assay is time consuming, low-throughput, and ineffective for measuring alcohol sensitivity in single flies.

Alternative ethanol sedation assays that do not require the elaborate inebriometer set-up allow for more convenient measurements but are still limited in throughput and generally require analyses of groups of flies rather than individuals[21,22,23,24,25]. Assessing single flies minimizes the potential for confounding effects due to group interactions, such as those stemming from social behaviors. Here, we present a simple, low-cost, high-throughput assay for assessing alcohol sedation sensitivity in large numbers of single flies.

# Protocol

### *1. Construction of the testing apparatus*

1. Create a cardboard template the size of a 24-well cell culture plate by tracing around the plate on cardboard and cutting out the designated area.
2. Cut a piece of small insect screen mesh the size of the cell culture plate using the cardboard template from step 1.1.
3. Prepare a 24-well cell culture plate by placing a small line of hot glue around the perimeter of the top of the plate using a hot glue gun and affixing the screen mesh on top of the open wells.
4. Secure a wooden craft stick to each of three sides of the same cell culture plate from step 1.3 using a hot glue gun. The modified cell culture plate should now resemble the plate diagram shown in Figure 2.1A and the experimental setup shown in Figure 2.2.

   NOTE: Prepare at least as many cell culture plates as will fit in the filming chambers (see below).

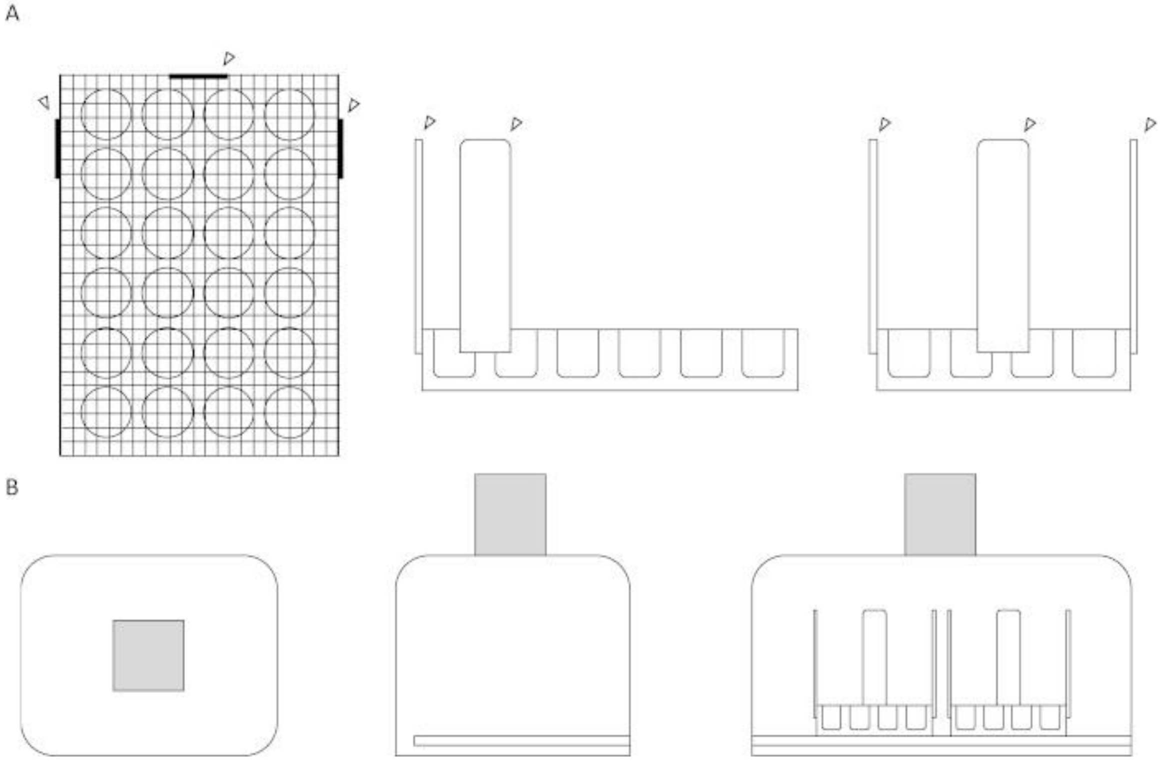


**Figure 2.1. Diagram of the testing apparatus and filming chamber.** (**A**) Upper Diagrams. The top, side, and front views of the testing apparatus are shown, respectively. A screen mesh lays flat on top of a 24-well cell culture plate. The wooden craft sticks, represented by the arrowheads, are attached to three adjacent sides for stability and alignment aid, two on the side of the well plate with six wells and one on the side of the plate with four wells. All attachments are hot glued onto the apparatus. (**B**) Lower Diagrams. The top, side, and front views of the assay set-up are shown, respectively. A slit is cut in the right side of the box, from the opening for the lid to the back of the opening, with the bottom of the slit level to the inner surface. The hole on the top of the box, the surface parallel to the ground, is centered for maximum video exposure. The shaded box represents the video camera.

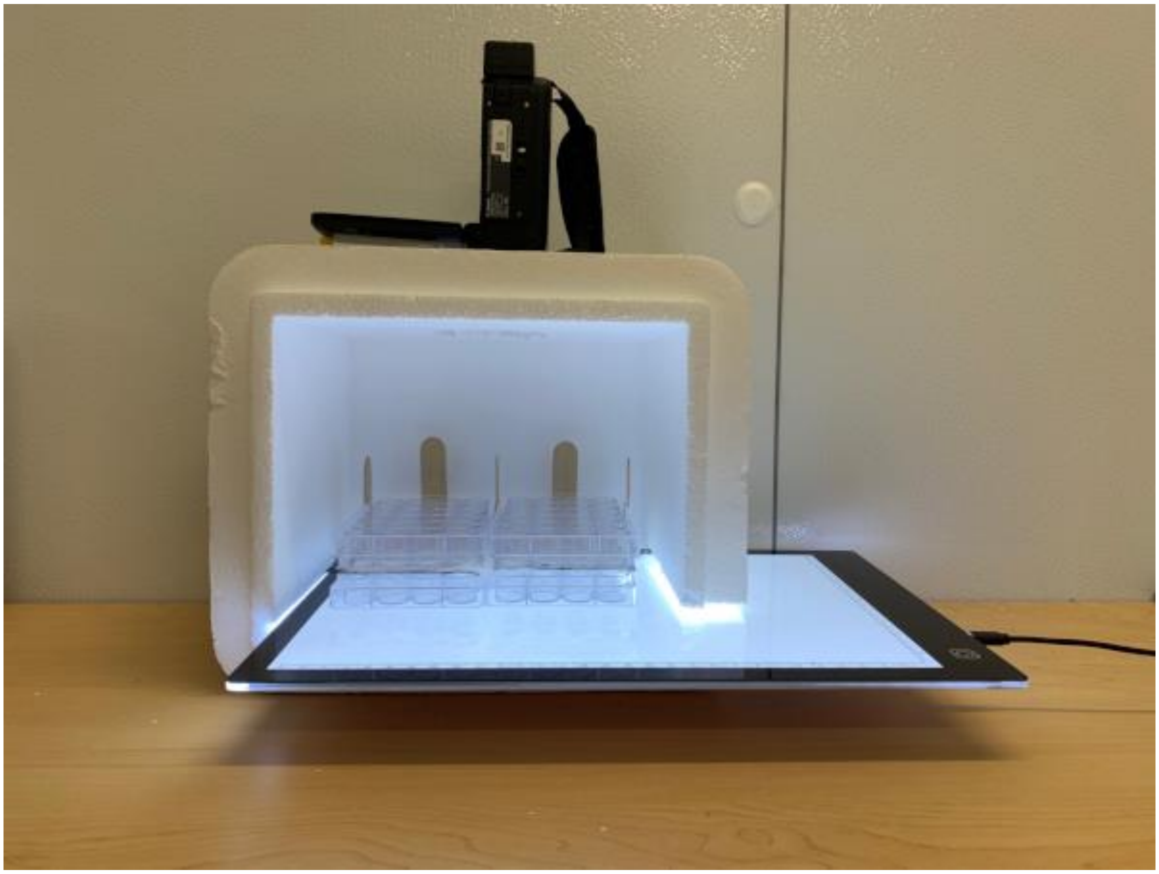


**Figure 2.2. Photograph of the assay system.** The video camera is placed on top of the polystyrene chamber, with the lens inserted in the cut-out hole, illustrated in the diagrams of Figure 2.1B. Two sets of modified 24-well cell culture plates rest on top of an illumination pad that is inserted in a slit through the side of the chamber.

### *2. Construction of the filming chamber*

1. Create a filming chamber by cutting a hole the size of the video camera lens on the side of a polystyrene box. Cut an additional slit the width of the illumination

pad in the opposite side of the polystyrene box. The filming chamber should resemble the filming chamber shown in Figure 2.1B and Figure 2.2.

2. Prepare the filming chamber for use by inserting the illumination pad into the slit and positioning the camera in the lens hole above the illumination pad.
3. Place all materials and perform all subsequent testing in a controlled environment, preferably a behavioral chamber with approximately 30% humidity, 25 °C temperature, uniform airflow, and noise levels less than 65 dB.

*3. Preparation of the testing apparatus and flies*

1. Pipette 1 mL of 100% ethanol through the screen mesh into each well.
2. Dry the screen mesh with a piece of cheesecloth.
3. Cut two pieces of cheesecloth the dimensions of the cell culture plate using the cardboard template created in step 1.1. Place them on top of the dry screen mesh of the modified cell culture plate containing ethanol from step 3.2.
4. Create a small piece of thin, flexible plastic cutting board by tracing around the cardboard template created in step 1.1 as a general guide and expanding the traced area by 1–2 cm on one of the short sides. Cut out the expanded traced area from the thin, flexible plastic cutting board. After cutting, ensure that the plastic still fits between the three wooden craft sticks on the testing apparatus, but hangs off one end by 1–2 cm.
5. (Optional) If an aspirator needs to be created, assemble an aspirator like the one shown in Figure 2.3 by first cutting a P1000 pipette tip in half. Insert the piece

with a larger diameter into one end of a ~30 cm piece of flexible tubing to serve as a mouthpiece.

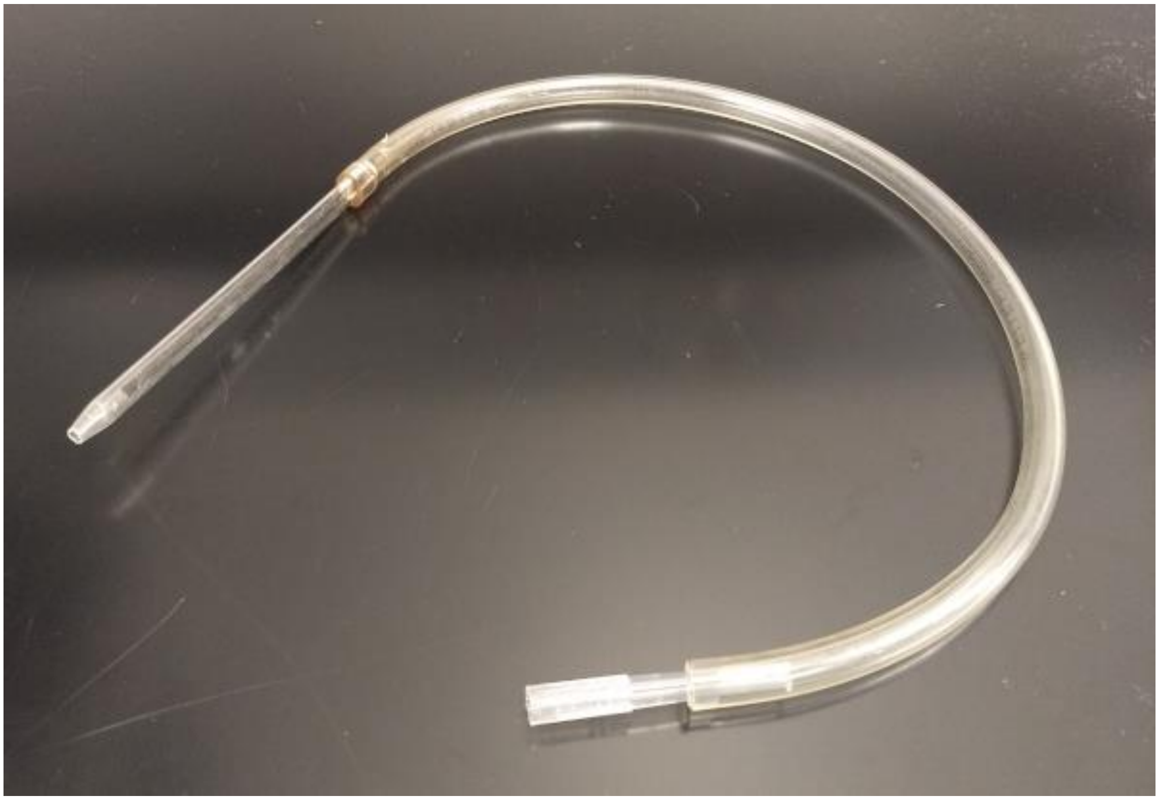


**Figure 2.3. A fly aspirator in which flies are collected with an interchangeable mouthpiece attached to flexible tubing and a wide bore serological pipette with a cotton gauze stopper**. The operator can aspirate a single fly into the pipette for transfer without anesthesia.

6. (Optional) To complete the aspirator assembly, cover the wide end of a 10 cm piece of serological pipette with gauze to prevent flies from getting into the tubing

and insert the pipette, gauze first, into the open end of the tubing to serve as a fly chamber. The aspirator should resemble that shown in Figure 2.3.

7. Using an aspirator (Figure 2.3, steps 3.5 and 3.6), aspirate one fly per well into a separate 24-well cell culture plate. Use the flexible plastic to cover any wells containing previously aspirated flies. Record the well position and any relevant genotype or phenotype information of each fly.
8. Hold the flexible plastic flush with the top of the cell culture plate containing the flies to prevent their escape and invert the plate onto the top of the modified cell culture plate with the ethanol. The sheet of flexible plastic should be resting on top of the sheets of cheesecloth. Align the inverted cell culture plate containing flies using the craft sticks to ensure each well with ethanol aligns with each well containing a fly.
9. The experimental setup should resemble Figure 2.2.

### *4. Testing the flies*

1. Ensure the illumination pad is lit at full brightness for maximum visual contrast. Start recording with the video camera.
2. To expose the flies to ethanol, carefully remove the plastic from between the well plate and testing apparatus, taking care not to dislodge the cheesecloth.
3. Terminate the video recording once all flies have lost postural control. Once it is suspected that all flies have lost postural control, tap firmly in the center of the plate to ensure that all flies have complete loss of postural control. If there is

movement, continue to record. Continue to tap periodically (every 1–2 min) until no movement occurs.

4. (Optional) To quickly recover the flies, remove only the top plate from the testing apparatus, revealing sedated flies resting on the cheesecloth. Aspirate individual flies into chosen containers for recovery.
5. Replace the ethanol in the modified cell culture plates with 1 mL of fresh 100% ethanol at least 1x every hour to control for evaporation and humidification of the ethanol and to maintain consistent ethanol exposure throughout the assay. Dry the screen mesh with cheesecloth.
6. Repeat for as many samples as desired.

   NOTE: For highest throughput, aspirate the next round of flies into new cell culture plates during the video recording. The protocol can be paused here, as the video recording can be reviewed later.

### *5. Determination of fly sedation time*

1. Record sedation time for each individual fly by watching the video recording. Sedation time is defined as the moment a fly loses complete postural control and locomotor ability. It is recommended to watch the film in reverse and record the time that the fly begins to move to ensure accuracy.

## Representative Results

Two 24-well microtiter plates could generate data simultaneously on 48 individual flies within as little as 10 min. Table 2.1 lists measurements of ethanol sedation times for 48 individual flies, males and females separately, of two DGRP lines with different sensitivities to alcohol exposure on development time and viability[13]. Flies of line RAL_555 were less sensitive than line RAL_177 (Figure 2.4, Table 2.2; $p < 0.0001$, ANOVA). Males and females of RAL_177 showed no sexually dimorphic effect (Figure 2.4, Table 2.2; $p > 0.1$, ANOVA), whereas females of line RAL_555 were less sensitive to ethanol exposure than the males (Figure 2.4, Table 2.2; $p < 0.006$, ANOVA). The large number of flies that can be measured simultaneously and the ability to measure sexes and different lines contemporaneously can increase accuracy by reducing error due to environmental variation.

**Table 2.1. Measurements of ethanol sedation times (s) of individual flies of (A) DGRP lines RAL_177 and (B) RAL_555 for separate sexes (n = 48).** See also Table 2.2, Figure 2.4.

| **A.** | **Ethanol Sedation Time (s)** | | | | | | **B.** | **Ethanol Sedation Time (s)** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Females** | | | **Males** | | | | **Females** | | | **Males** | | |
| | 414 | 365 | 477 | 423 | 568 | 309 | | 937 | 742 | 622 | 460 | 331 | 498 |
| | 201 | 384 | 498 | 411 | 523 | 626 | | 791 | 619 | 197 | 467 | 455 | 562 |
| | 228 | 364 | 333 | 440 | 403 | 267 | | 504 | 744 | 513 | 570 | 582 | 506 |
| | 440 | 416 | 404 | 408 | 422 | 384 | | 970 | 540 | 369 | 865 | 533 | 492 |
| | 888 | 283 | 285 | 322 | 369 | 287 | | 595 | 550 | 606 | 392 | 544 | 345 |
| | 1079 | 519 | 315 | 393 | 376 | 284 | | 418 | 709 | 553 | 308 | 477 | 388 |
| | 718 | 287 | 432 | 275 | 206 | 411 | | 366 | 564 | 558 | 385 | 576 | 377 |
| | 598 | 337 | 398 | 279 | 631 | 372 | | 437 | 692 | 578 | 460 | 511 | 412 |
| | 241 | 398 | 364 | 347 | 374 | 808 | | 665 | 729 | 484 | 532 | 425 | 354 |
| | 229 | 423 | 534 | 386 | 396 | 628 | | 312 | 576 | 305 | 334 | 531 | 506 |
| | 388 | 488 | 451 | 523 | 322 | 533 | | 682 | 638 | 420 | 560 | 548 | 379 |
| | 252 | 529 | 375 | 427 | 330 | 540 | | 1045 | 741 | 708 | 832 | 509 | 472 |
| | 674 | 401 | 303 | 401 | 307 | 311 | | 394 | 675 | 381 | 477 | 449 | 784 |
| | 303 | 453 | 351 | 429 | 525 | 262 | | 540 | 690 | 520 | 556 | 495 | 226 |
| | 258 | 483 | 302 | 389 | 562 | 319 | | 356 | 615 | 336 | 454 | 524 | 590 |
| | 346 | 426 | 385 | 416 | 596 | 287 | | 626 | 678 | 840 | 634 | 677 | 509 |

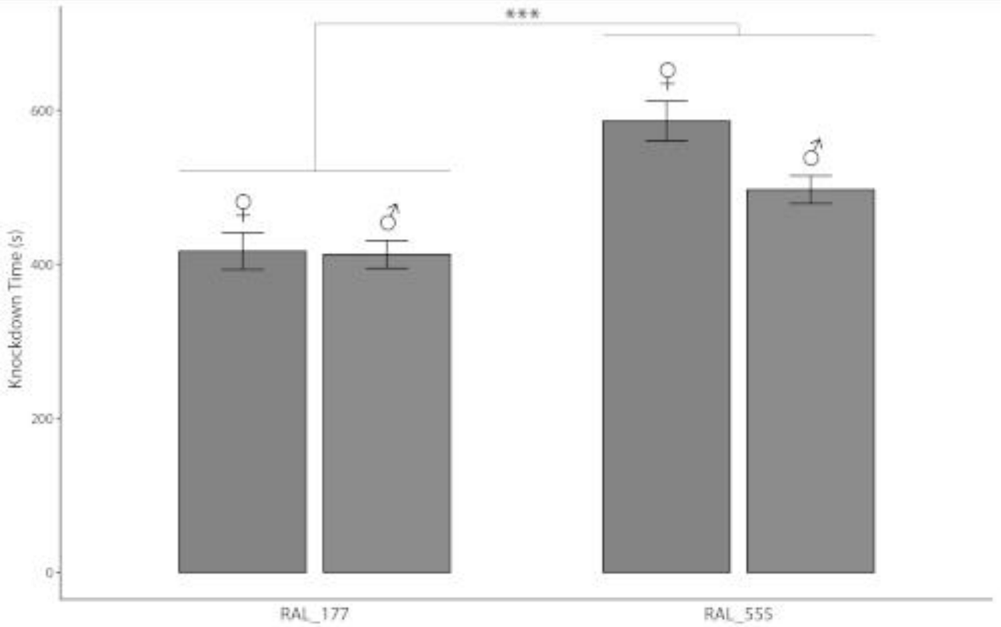


**Figure 2.4. Alcohol sedation times of DGRP lines RAL_177 and RAL_555.** The bars represent means and the error bars SEM (n = 48). Sedation times for RAL_177 flies were less than those for RAL_55 flies ($p < 0.0001$, ANOVA). Individual data points are indicated in Table 2.1. Additional statistically significant differences between sexes and lines are indicated in the text and in Table 2.2.

**Table 2.2. Analyses of variance for sedation time across sex and DGRP line.** The model used was $Y = \mu + L + S + LxS + \varepsilon$, where $\mu$ is the overall mean, $L$ is the fixed effect of the DGRP line (RAL_177, RAL_555), $S$ is the fixed effect of sex (male, female), $LxS$ is the interaction term (fixed), and $\varepsilon$ is the error term. The models $Y = \mu + L + \varepsilon$ and $Y = \mu + S + \varepsilon$ were used for the reduced models. Line, Sex, and the Line x Sex interaction term were all significant in the full model at $\alpha < 0.05$. Reduced models by sex and DGRP line RAL_555 were also significant at $\alpha < 0.01$. See also Table 2.1, Figure 2.4. df = degrees of freedom, SS = Type I Sums of Squares.

| **Analysis** | **Source of Variation** | **df** | **SS** | **F-Value** | **P-value** |
|---|---|---|---|---|---|
| Full Model Pooled | Line | 1 | 769627 | 34.869 | <0.0001 |
| | Sex | 1 | 105001 | 4.757 | 0.0304 |
| | Line x Sex | 1 | 86021 | 3.897 | 0.0498 |
| | Error | 188 | 4149491 | | |
| Reduced Model Females | Line | 1 | 685126 | 23.58 | <0.0001 |
| | Error | 94 | 2730718 | | |
| Reduced Model Males | Line | 1 | 170522 | 11.3 | 0.0011 |
| | Error | 94 | 1418774 | | |
| Reduced Model RAL_177 | Sex | 1 | 473 | 0.023 | 0.8800 |
| | Error | 94 | 1943741 | | |
| Reduced Model RAL_555 | Sex | 1 | 190549 | 8.12 | 0.0054 |
| | Error | 94 | 2205751 | | |

## Discussion

Here, we present a simple, inexpensive, and high-throughput method for assessing sedation time due to ethanol exposure in *Drosophila melanogaster*. Unlike many current methods, which require group analyses, this assay enables a single person to collect individual sedation time data for ~2,000 flies within an 8 h work period. We found that a single person can score 48 flies for sedation time in about 5 min. At this rate, 2,000 flies can be scored in approximately 4 h, though scoring can be conducted later. With our assay, the recorded sedation time for most flies ranges from 5–15 min at an exposure to 1 mL of 100% ethanol. Lower concentrations of ethanol or smaller delivery volumes will result in longer sedation times.

Current methods for assessing sedation time require testing large numbers of flies without readily enabling measurements on single individuals[15,16,17,18,19,20,21,22,23,24,25,26]. Many current sedation and sensitivity assays rely upon ST50[22,23,24], the timepoint at which 50% of the flies are sedated as a result of ethanol exposure. Although obtaining the ST50 for groups of flies was not the primary motivation for developing this assay, the video recordings demonstrate higher utility compared to current methods, as the recordings can be used to ascertain the ST50 for groups of individually tested flies and to measure the percentage of flies that satisfy a given criterion (e.g., loss of postural control) at any time point. It should be noted that such video analyses would require additional time.

Unlike current inebriometer assays, the method we describe does not require specialized tools to set up and can be performed in any laboratory using common materials. Using this method, we have obtained reliable and consistent sedation times for individual flies. The assay can, in principle, be extended to assess the effects of exposure to any volatile substance. The assay can also be applied to measure effects of acute toxicity of volatiles on other insects, including other fly species. Individual sedation time data can be used to assess the extent of phenotypic variation within a population, such as the DGRP.

We used small insect screen mesh to prevent direct contact with the ethanol solution while allowing adequate quantities of ethanol vapors to reach the fly. The layer of white cheesecloth on top of the screen mesh provides visual contrast between the fly and the surface below and ensures that flies do not get caught in the screen mesh, which could lead to ambiguous determination of loss of postural control. Commercially available membranes that are porous to water and air gave inconsistent results and were insufficiently penetrable to ethanol vapors. We intentionally used small insect screen mesh because it is a uniformly porous material that minimizes variation in ethanol exposure as a result of fly position within a well. Modifications can be made to this protocol based on available materials, although we recommend a controlled behavioral chamber, access to 90%–100% ethanol close to the fly, and uniform ethanol exposure. Fly position within the cell culture plates should be randomized between replicates to avoid positional bias. For larger experiments that require use of this assay across multiple days and are therefore subject to environmental variation that could influence assay results (e.g., changes in barometric pressure)[27], we strongly recommend that flies be tested at the same time each day and randomized both within and across days, especially if different lines and/or sexes are to be compared against one another.

The method we developed is best suited for measuring the effect of acute alcohol exposure but is not suitable for obtaining consumption data or modeling addiction. Alcohol sedation sensitivity data obtained from this assay can, however, be integrated with other measures of alcohol-related phenotypes. One limitation of the system is that the vertical height of standard cell culture plates allows for vertical fly movement that cannot be readily tracked by video for detailed assessment of overall activity or locomotion. However, this limitation does not affect accurate assessment of sedation time. When using flies of different genotypes (e.g., in DGRP-derived outbred populations[28]), this assay also enables retrieval of individual flies to collect pools of flies with contrasting phenotypes for bulk DNA sequencing and extreme QTL mapping[29,30]. Overall, this assay permits rapid, inexpensive collection of alcohol sedation data on large numbers of single flies.

## Acknowledgements

This work was supported by grants DA041613 and GM128974 from the National Institutes of Health to TFCM and RRHA.

# References

1. Heath, A. C., et al. Genetic differences in alcohol sensitivity and the inheritance of alcoholism risk. *Psychological Medicine*. **29** (5), 1069-1081 (1999).

2. Schuckit, M. A., Smith, T. L. The relationships of a family history of alcohol dependence, a low level of response to alcohol and six domains of life functioning to the development of alcohol use disorders. *Journal of Studies on Alcohol*. **61** (6), 827-835 (2000).

3. Trim, R. S., Schuckit, M. A., Smith, T. L. The relationships of the level of response to alcohol and additional characteristics to alcohol use disorders across adulthood: a discrete-time survival analysis. *Alcoholism: Clinical and Experimental Research*. **33** (9), 1562-1570 (2009).

4. Schuckit, M. A., Smith, T. L. Onset and course of alcoholism over 25 years in middle class men. *Drug and Alcohol Dependence*. **113** (1), 21-28 (2011).

5. Morozova, T. V., Mackay, T. F. C., Anholt, R. R. H. Genetics and genomics of alcohol sensitivity. *Molecular Genetics and Genomics*. **289** (3), 253-269 (2014).

6. Heberlein, U., Wolf, F. W., Rothenfluh, A., Guarnieri, D. J. Molecular genetic analysis of ethanol intoxication in Drosophila melanogaster. *Integrative and Comparative Biology*. **44** (4), 269-274 (2004).

7. Engel, G. L., Taber, K., Vinton, E., Crocker, A. J. Studying alcohol use disorder using Drosophila melanogaster in the era of 'Big Data'. *Behavioral and Brain Functions*. **15** (1), 7 (2019).

8. Singh, C. M., Heberlein, U. Genetic control of acute ethanol-induced behaviors in Drosophila. *Alcoholism: Clinical and Experimental Research*. **24** (8), 1127-1136 (2000).

9. Mackay, T. F. C., et al. The Drosophila melanogaster Genetic Reference Panel. *Nature*. **482** (7384), 173-178 (2012).

10. Huang, W., et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Research*. **24** (7), 1193-1208 (2014).

11. Morozova, T. V., et al. A Cyclin E centered genetic network contributes to alcohol-induced variation in Drosophila development. *G3*. **8** (8), 2643-2653 (2018).

12. Morozova, T. V., et al. Polymorphisms in early neurodevelopmental genes affect natural variation in alcohol sensitivity in adult drosophila. *BMC Genomics*. **16**, 865 (2015).

13. Scholz, H., Ramond, J., Singh, C. M., Heberlein, U. Functional ethanol tolerance in Drosophila. *Neuron*. **28** (1), 261-271 (2000).

14. Berger, K. H., Heberlein, U., Moore, M. S. Rapid and chronic: two distinct forms of ethanol tolerance in Drosophila. *Alcoholism: Clinical and Experimental Research*. **28** (10), 1469-1480 (2004).

15. Morozova, T. V., Anholt, R. R. H., Mackay, T. F. C. Transcriptional response to alcohol exposure in Drosophila melanogaster. *Genome Biology*. **7** (10), 95 (2006).

16. Weber, K. E. An apparatus for measurement of resistance to gas-phase agents. *Drosophila Information Service*. **67**, 91-93 (1988).

17. Weber, K. E., Diggins, L. T. Increased selection response in larger populations. II. Selection for ethanol vapor resistance in Drosophila melanogaster at two population sizes. *Genetics*. **125** (3), 585-597 (1990).

18. Cohan, F. M., Graf, J. D. Latitudinal cline in Drosophila melanogaster for knockdown resistance to ethanol fumes and for rates of response to selection for further resistance. *Evolution*. **39** (2), 278-293 (1985).

19. Scholz, H., Franz, M., Heberlein, U. The hangover gene defines a stress pathway required for ethanol tolerance development. *Nature*. **436** (7052), 845-847 (2005).

20. Morozova, T. V., Anholt, R. R. H., Mackay, T. F. C. Phenotypic and transcriptional response to selection for alcohol sensitivity in Drosophila melanogaster. *Genome Biology*. **8** (10), 231 (2007).

21. Morozova, T. V., et al. Alcohol sensitivity in Drosophila: Translational potential of systems genetics. *Genetics*. **83**, 733-745 (2009).

22. Bhandari, P., Kendler, K. S., Bettinger, J. C., Davies, A. G., Grotewiel, M. An assay for evoked locomotor behavior in Drosophila reveals a role for integrins in ethanol sensitivity and rapid ethanol tolerance. *Alcoholism: Clinical and Experimental Research*. **33** (10), 1794-1805 (2009).

23. Sandhu, S., Kollah, A. P., Lewellyn, L., Chan, R. F., Grotewiel, M. An inexpensive, scalable behavioral assay for measuring ethanol sedation sensitivity and rapid tolerance in Drosophila. *Journal of Visualized Experiments*. (98), e52676 (2015).

24. Urizar, N. L., Yang, Z., Edenberg, H. J., Davis, R. L. Drosophila homer is required in a small set of neurons including the ellipsoid body for normal ethanol sensitivity and tolerance. *The Journal of Neuroscience*. **27** (17), 4541-4551 (2007).

25. Wolf, F. W., Rodan, A. R., Tsai, L. T., Heberlein, U. High-resolution analysis of ethanol-induced locomotor stimulation in Drosophila. *The Journal of Neuroscience*. **22** (24), 11035-11044 (2002).

26. Cohan, F. M., Hoffmann, A. A. Genetic divergence under uniform selection. II. Different responses to selection for knockdown resistance to ethanol among Drosophila melanogaster populations and their replicate lines. *Genetics*. **114** (1), 145-164 (1986).

27. Pohl, J. B., et al. Circadian genes differentially affect tolerance to ethanol in Drosophila. *Alcoholism: Clinical and Experimental Research*. **37** (11), 1862-1871 (2013).

28. Huang, W., et al. Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proceedings of the National Academy of Sciences of the United States of America*. **109**, 15553-15559 (2012).

29. Ehrenreich, I. M., et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature*. **464** (7291), 1039-1042 (2010).

30. Anholt, R. R. H., Mackay, T. F. C. The road less traveled: From genotype to phenotype in flies and humans. *Mammalian Genome*. **29**, 5-23 (2018).

# CHAPTER THREE

# DEVELOPMENTAL ALCOHOL EXPOSURE IN DROSOPHILA: EFFECTS ON ADULT PHENOTYPES AND GENE EXPRESSION IN THE BRAIN

Mokashi SS*, Shankar V*, MacPherson RA, Hannah RC, Mackay TFC, Anholt RRH. Developmental Alcohol Exposure in Drosophila: Effects on Adult Phenotypes and Gene Expression in the Brain. *Front Psychiatry*. 2021 Jul 22;12:699033. doi: 10.3389/fpsyt.2021.699033. PMID: 34366927; PMCID: PMC8341641.

*Equal contribution

## Author Contribution Statement

My contributions to the work described in this chapter are as follows: I was involved in experimental design and Drosophila behavioral and viability phenotyping alongside Dr. Sneha Mokashi and Rachel Hannah. I performed data analysis for the behavioral and viability phenotypes. I was also involved in biological interpretation of the sequencing data and writing of the manuscript.

## Introduction

Prenatal exposure to ethanol can trigger a wide range of adverse physiological, behavioral, and cognitive outcomes, collectively termed fetal alcohol spectrum disorder (FASD) (1–4). Fetal alcohol syndrome (FAS) has the most severe manifestations of all

FASDs, including craniofacial dysmorphologies, neurocognitive deficiencies, and behavioral disorders such as hyperactivity, attention deficit disorder and motor coordination anomalies (1, 5–7). FAS/FASD is the most common preventable pediatric disorder, often diagnostically confounded with autism spectrum disorder (8). Time, dose, and frequency of exposure are often unknown, and manifestations of FASD are diverse and become evident long after exposure. The Centers for Disease Control and Prevention found that 1 in 10 pregnant women report alcohol use and more than 3 million women in the USA are at risk of exposing their developing fetus to alcohol, despite warning labels on alcoholic beverages that indicate possible effects on prenatal development (9). Adverse consequences of fetal alcohol exposure extend throughout the lifespan.

Determining the effects of developmental alcohol exposure on adult phenotypes and gene expression in the adult brain is challenging in human populations, but can be addressed in model organisms. *Drosophila melanogaster* is an excellent model to study developmental effects of alcohol exposure, as we can control the genetic background and environmental conditions for large numbers of individuals without regulatory restrictions and at low cost. Importantly, flies exposed to alcohol experience loss of postural control, sedation, and development of tolerance (10–13), resembling human alcohol intoxication. Previous studies on the effects of developmental alcohol exposure in Drosophila showed reduced viability and delayed development time (14, 15), reduced adult body size (14) and disruption of neural development (16). Developmental exposure to alcohol was associated with reduction in the expression of a subset of insulin-like peptides and the insulin receptor (14), dysregulation of lipid metabolism and concomitant increased oxidative stress (17), and reduced larval food intake due to altered neuropeptide F signaling (18).

Here, we show that developmental alcohol exposure in Drosophila results in decreased viability, reduced sensitivity to ethanol, and disrupted sleep and activity patterns. Single cell RNA sequencing on adult fly brains following developmental alcohol exposure shows widespread sexually dimorphic changes in gene expression. These changes in gene expression resemble changes observed previously following cocaine exposure (19), indicating common neuronal and glial elements that respond to alcohol and cocaine consumption.

## Methods

### *Drosophila Stocks and Exposure to Ethanol*

*Drosophila melanogaster* of the wild type Canton S (B) strain were maintained on cornmeal/yeast/molasses-agar medium supplemented with yeast at 25°C on a 12 h light: dark cycle with 50% humidity, in controlled adult density vials to prevent overcrowding. We allowed 5 males and 5 females to mate for 2 days and aged their progeny for 3–5 days after eclosion. We then placed 50 males and 50 females into large egg collection cages on grape juice agar and yeast paste. We acclimatized the flies to the cages for 24 h with grape juice plate changes every 12 h, and collected up to 12-h old eggs with a blunt metal needle. We placed the eggs on cornmeal-agar-molasses medium (control) or on cornmeal-agar-molasses medium containing 10% (v/v) ethanol (ethanol) without yeast. We collected 50 eggs per vial and set up 10–15 vials per condition per collection week over a 48-h period (Figure 3.1). After eclosion, flies were transferred to control medium without yeast and aged as indicated for the relevant experiments. Unless otherwise indicated, all behavioral assays were performed in a controlled environment at 25°C.

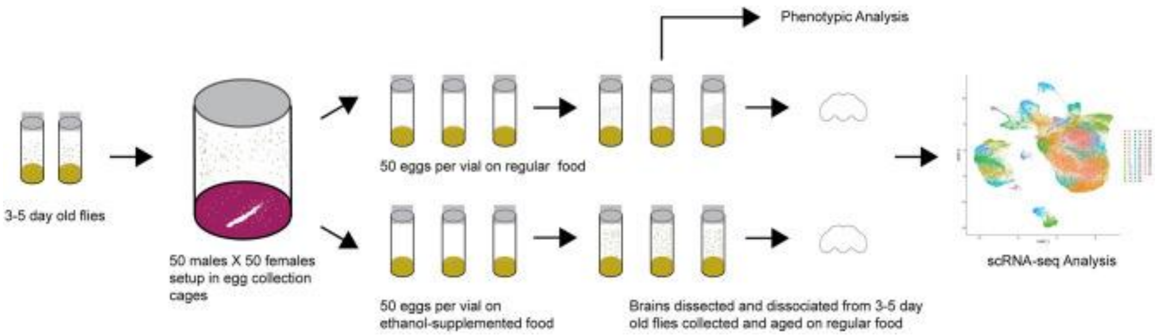

Phenotypic Analysis
3-5 day old flies
50 males X 50 females setup in egg collection cages
50 eggs per vial on regular food
50 eggs per vial on ethanol-supplemented food
Brains dissected and dissociated from 3-5 day old flies collected and aged on regular food
scRNA-seq Analysis

**Figure 3.1 Diagram of the experimental design**

### *Viability*

The number of flies that emerged from vials into which 50 eggs had been placed were counted and the data were analyzed using the “PROC GLM” command (Type III) in SAS v3.8 (Cary, NC) according to the model $Y = \mu + T + \varepsilon$, where $Y$ is the number of eclosed flies, $\mu$ is the population mean, $T$ is the fixed effect of treatment (flies reared on control or ethanol medium), and $\varepsilon$ is the residual error.

### *Ethanol Sensitivity*

We measured ethanol sedation time as described previously (20) on 44–48 3–5 day old flies per sex per treatment. Ethanol sedation time was assessed between 8:30 a.m. and 11:30 a.m. The number of seconds required for flies to lose postural control was analyzed using the "PROC GLM" command (Type III) in SAS v3.8 according to the model $Y = \mu + T + S + TxS + \varepsilon$, where $Y$ is the time to sedation, $\mu$ is the population mean, $T$ is the fixed effect of treatment (control or ethanol medium), $S$ is the fixed effect of sex, *a*nd $\varepsilon$ is the residual error.

### *Sleep and Activity*

Flies reared on either control or ethanol medium were placed in Drosophila Activity Monitors (DAM) (TriKinetics, Waltham, MA) containing a 5% sucrose, 2% agar medium at 1–2 days of age, and monitored for 7 days on a 12 h light-dark cycle. Activity was recorded as counts every time the fly interrupts an infrared beam. Sleep was defined as at least 5 min of inactivity. Only data from flies that survived the entire testing period were included, resulting in 57–64 flies per sex per treatment for analysis. Raw DAM monitor data were run in ShinyR-DAM (21), and the outputs were downloaded and parsed according to phenotype (e.g., day/night, sleep/activity, bout length/bout count) for subsequent statistical analyses. The data were analyzed using the "PROC MIXED" command (Type III) in SAS v3.8 according to the model $Y = \mu$
$+ T + S + TxS + Rep(TxS) + \varepsilon$, where $Y$ is the sleep or activity phenotype, $\mu$ is the population mean, $T$ is the fixed effect of treatment (control or ethanol medium), $S$ is the

fixed effect of sex, *Rep* is the random effect of replicate and ε is the residual error. Reduced models were also performed for each sex.

### *Brain Dissociation and Single Cell RNA Sequencing*

For single cell RNA sequencing, we collected duplicate samples of 20 brains for each sex from flies reared on control or ethanol medium. We dissociated the brains as previously described after incubation with 450 μl of collagenase solution [50 ul of fresh 25 mg/ml collagenase (Gibco) in sterile water + 400 μl of Schneider's medium] for 30 min followed by stepwise trituration - P200 pipette 5 times, 23G needle pre-wetted with PBS + BSA 5 times, and 27G pre-wetted needle 5 times (19). The resulting suspension was passed through a pre-wetted 10 μm strainer (Celltrics, Görlitz, Germany) with gentle tapping. We counted live cells using a hemocytometer with trypan blue exclusion and proceeded with GEM generation using the Chromium controller (10X Genomics, Pleasanton, CA) for samples with >500 live cells/μl. We prepared libraries in accordance with 10X Genomics v3.1 protocols. We determined fragment sizes using Agilent Tapestation kits (Agilent, Santa Clara, CA)—d5000 for amplified cDNA and d1000 for libraries. We measured the concentrations of amplified cDNA and final libraries using a Qubit 1X dsDNA HS kit (Invitrogen, Waltham, MA) and a qPCR based library quantification kit (KAPA Biosystems, Roche, Basel, Switzerland). We used 12 cycles for the cDNA amplification and 12 cycles for indexing PCR. We sequenced the final libraries on an Illumina NovaSeq6000.

### *Single Cell RNA Sequencing Data Analysis and Bioinformatics*

We used the *mkfastq* pipeline within Cell Ranger v3.1 (10X Genomics, Pleasanton, CA) to convert BCL files from the sequence run folder to demultiplexed FASTQ files. We used the *mkref* pipeline to index the release 6 version of the *D. melanogaster* reference *GCA_000001215.4* from NCBI Genbank. For alignment, we used the *count* pipeline within Cell Ranger v3.1 with the expected cell count parameter set to 5,000 cells. We imported raw expression counts output for each sample from the Cell Ranger pipeline and analyzed these data using the Seurat v3 package in R (22). We normalized counts by regularized negative binomial regression using the *scTransform* pipeline (23). We performed integration of samples using the *SCT* method. *RunUMAP* and *FindNeighbors* functions were used with 10 dimensions to ordinate expression space and reduce data dimensionality. To identify cell-type clusters, we used unsupervised clustering using the *FindClusters* function and assigned the origin of clustered cells based on well-established biomarkers.

We used the Pearson residuals output from the *scTransform* pipeline as input for differential expression calculation (23). We used the *MAST* algorithm as the testing methodology in the *FindMarkers* function for each cluster to calculate differential expression, which allows for the incorporation of the cellular detection rate, defined as a fraction of genes expressed in each cell, as a covariate (24). *P*-values for differential expression were adjusted for multiple-hypothesis testing using a Bonferroni correction, and adjusted *p*-values that are <0.05 were considered statistically significant.

Interaction networks were produced using the unique list of differentially expressed genes aggregated from all clusters and the stringApp (25) within Cytoscape (26).

The code for all analyses can be found here: https://github.com/vshanka23/The-Drosophila-Brain-after-developmental-ethanol-exposure-at-Single-Cell-Resolution/blob/main/Rcode_for_analysis.R

### *Data Availability Statement*

The datasets for this study can be found in the GEO repository under accession number GSE172231.

## Results

### *Effects of Developmental Alcohol Exposure on Adult Phenotypes*

Exposure of flies to ethanol during the embryonic and larval stages resulted in an 8.9% reduction in viability compared to flies reared on control medium (Figure 3.2A). The adult flies exposed to ethanol during development did not show any overt morphological abnormalities. We next asked whether developmental alcohol exposure would alter sensitivity to acute alcohol exposure as adults. We reared developing flies on ethanol medium and transferred the adults to control medium immediately after eclosion. The flies that developed on ethanol medium showed reduced sensitivity (longer sedation times) to acute alcohol exposure in both sexes, indicating increased tolerance to acute alcohol exposure compared to flies that developed on control medium (Figure 3.2B).

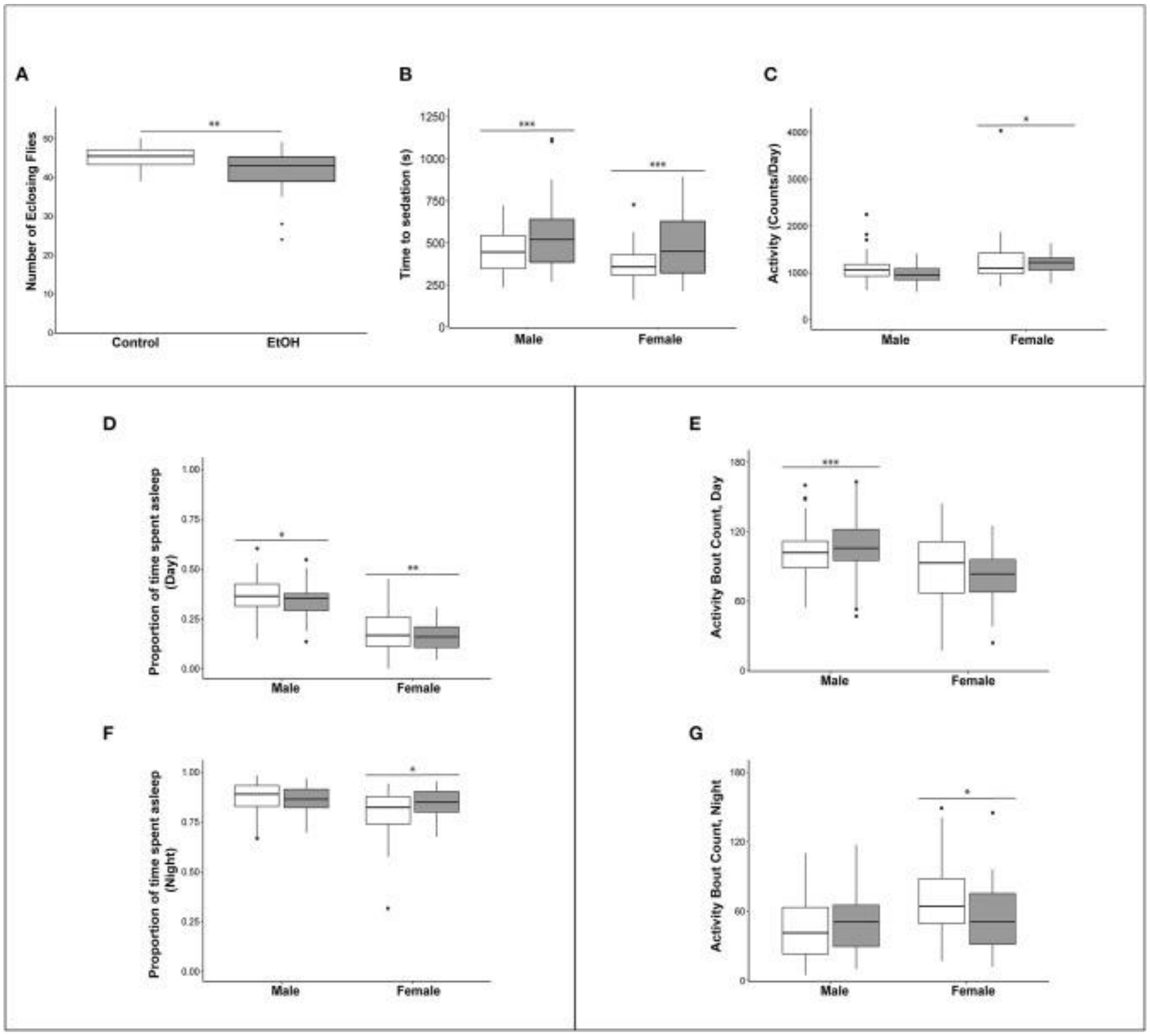


**Figure 3.2. Effects of developmental alcohol exposure on viability and behavioral phenotypes in adult flies. (A)** Boxplots of viability (*n* = 12 reps of 50 embryos per treatment), **(B)** Ethanol sensitivity (*n* = 43–49, 3–5 day old flies per sex per treatment), **(C)** Activity, **(D)** Proportion of daytime sleep, **(E)** Activity bouts during the day. **(F)** Proportion of night time sleep, **(G)** Activity bouts during the night. Day hours

are from 7 a.m. to 7 p.m., lights on 7 h after hour zero. Gray boxes indicate flies reared on medium supplemented with 10% (v/v) ethanol and white boxes indicate control flies grown on regular medium. $n$ = 57–64 flies per sex per treatment for all sleep and activity phenotypes. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Actograms are shown in Supplementary Figure S3.1.

Children with FASD often have disturbed sleep (27, 28). Therefore, we used the Drosophila Activity Monitor system to assess the effects of developmental alcohol exposure on adult activity and sleep patterns and found that exposure to alcohol during development had sex-specific effects on these phenotypes. Overall activity in males was not affected by the ethanol treatment, but females exposed to ethanol were more active (Figure 3.2C and Supplementary Table S3.1). Ethanol exposure reduced sleep during the day in both sexes (Figure 3.2D), and day sleep in males was fragmented, with an increase in activity bouts (Figure 3.2E). In contrast, females compensated for increased activity and reduced daytime sleep with extended periods of night sleep (Figure 3.2F) with a reduced number of activity bouts (Figure 3.2G and Supplementary Figure S3.1; Supplementary Table S3.1).

### *Effects of Developmental Alcohol Exposure on Gene Expression in the Brain*

We performed single cell RNA sequencing to assess the effects of developmental alcohol exposure on gene expression in the brain in males and females, with two replicates per sex and treatment (Figure 3.1). We obtained a total of 108,571 cells across all samples, which corresponds to ~10% of all cells in a Drosophila brain (Supplementary Table S3.2). We visualized these data using the Uniform Manifold Approximation and Projection (UMAP) non-linear dimensionality reduction method (29), which showed that all samples were uniformly represented (Figure 3.3 and Supplementary Table S3.2). Unsupervised clustering of the dataset generated 43 cell clusters, which represent the major regions of the Drosophila brain, including neuronal and glial populations, and all major neurotransmitter cell types (Figure 3.4 and Supplementary Table S3.3). We identified seven distinct populations of GABAergic neurons, two subpopulations of Kenyon cells of the mushroom bodies (integrative centers for experience-dependent modulation of behavior), and several distinct populations of glia, including two separate clusters of astrocytes as well as surface glia that form the blood-brain barrier (Figure 3.4).

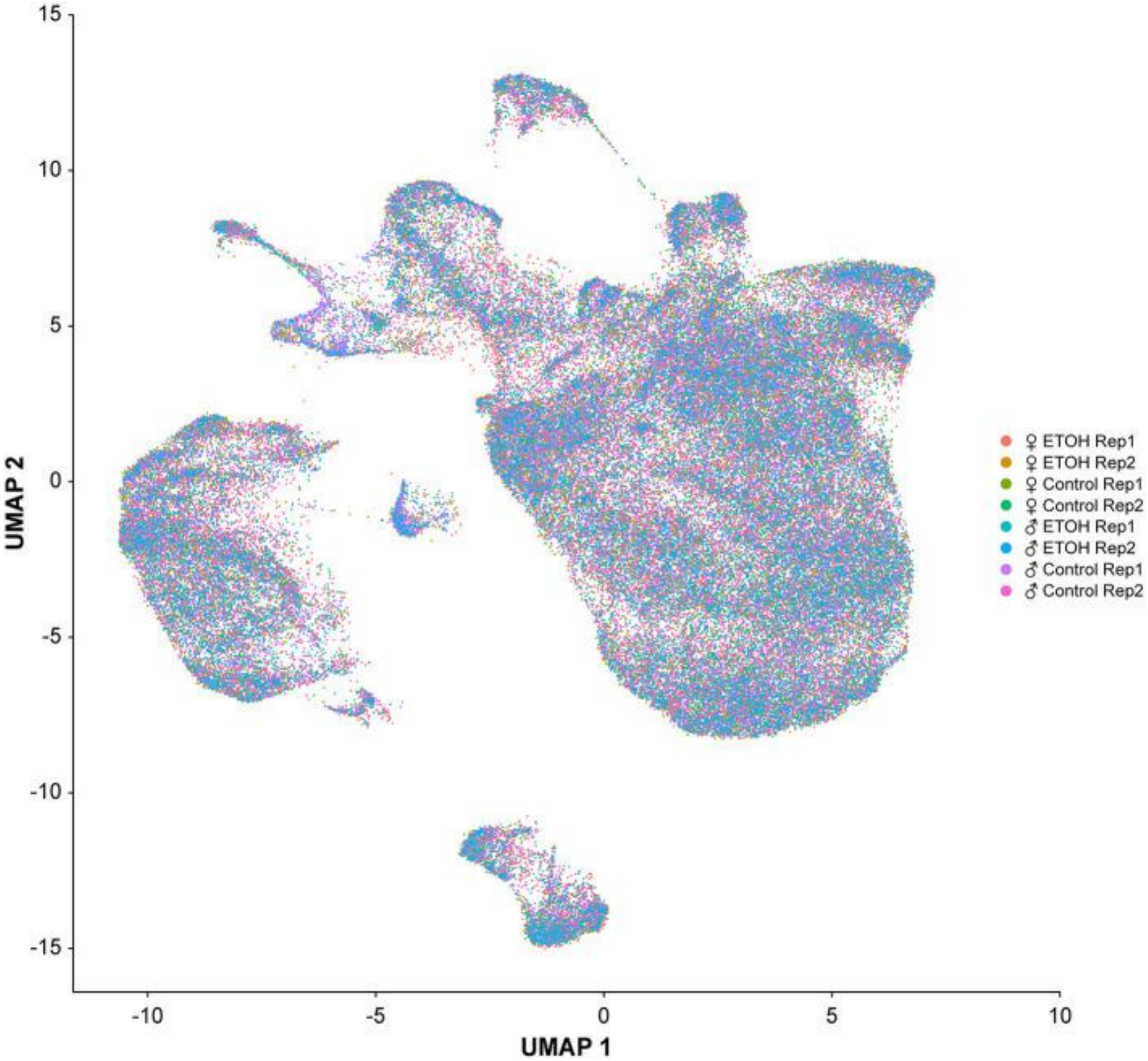


**Figure 3.3. Uniformity across samples of single cell transcriptomes.** Gene expression patterns of single cells (*n* = 108, 571) from all eight samples are represented in low dimensional space using a graph-based, non-linear dimensionality reduction method (UMAP). Individual dots represent the transcriptome of each cell and the colors of the dots represent the samples to which the cells belong.

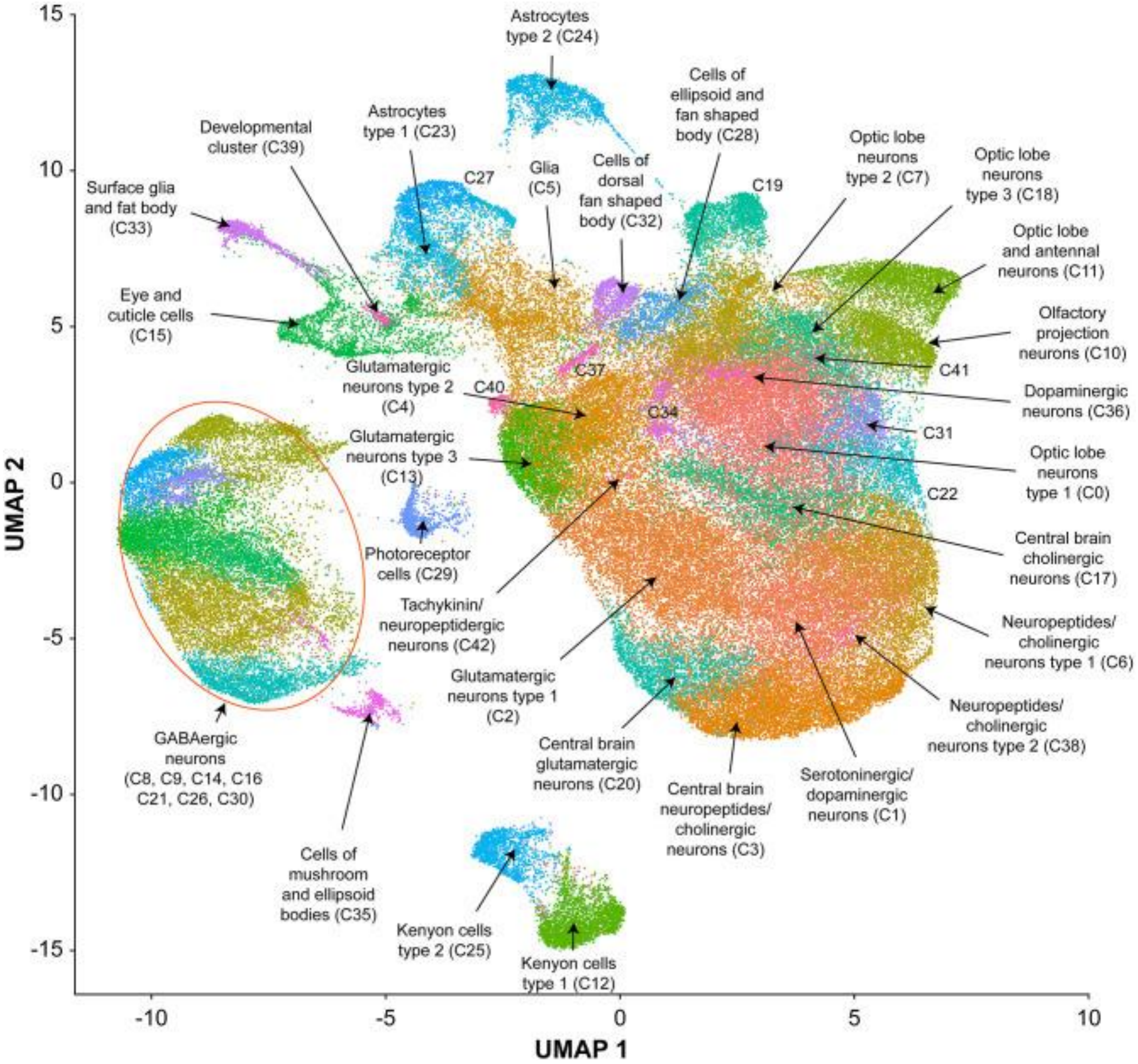


**Figure 3.4. UMAP visualization and annotation of cell clusters.** Cells were clustered based on their expression pattern using the unsupervised shared nearest neighbor (SNN) clustering algorithm. Individual dots represent each cell and the colors of the dots represent the cluster to which the cells belong. Annotation of cell types from clusters was performed by cross-referencing cluster-defining genes across FlyBase (30) and published literature (Supplementary Table S3.3).

We combined all differentially expressed genes from all clusters and performed differential expression analyses. We found 119 transcripts in males and 148 transcripts in females with altered abundances after developmental alcohol exposure at a Bonferroni adjusted $p < 0.05$. We identified 61 upregulated and 25 downregulated genes in males, and 57 upregulated and 34 downregulated genes in females at a threshold of $|\log_e FC| > 0.25$ (Figure 3.5 and Supplementary Tables S3.4, S3.5). Increasing the stringency to $|\log_e FC| > 1.0$ (Bonferroni adjusted $p < 0.05$) retained 36 upregulated and 10 downregulated genes in males and 32 upregulated and 20 downregulated genes in females (Supplementary Figure S3.2). Differential expression patterns are sexually dimorphic, as observed previously for cocaine-induced modulation of gene expression (19), with only 32 differentially expressed genes in common between the sexes. Changes in gene expression in the mushroom bodies, represented by cluster C12, are primarily observed in females. Developmental alcohol exposure modulates expression of several genes in glia, represented by clusters C5, C15, C23, C24, and C33, in a sexually dimorphic pattern (Figure 3.5). Especially noteworthy is the prominent differential expression of *lncRNA:CR31451*, a long non-coding RNA of unknown function, in multiple neuronal populations. This transcript is globally upregulated in males but downregulated in females (Figure 3.5 and Supplementary Figure S3.2). Among all differentially expressed genes, ~58% have human orthologs (DIOPT score ≥ 3; Supplementary Table S3.6).

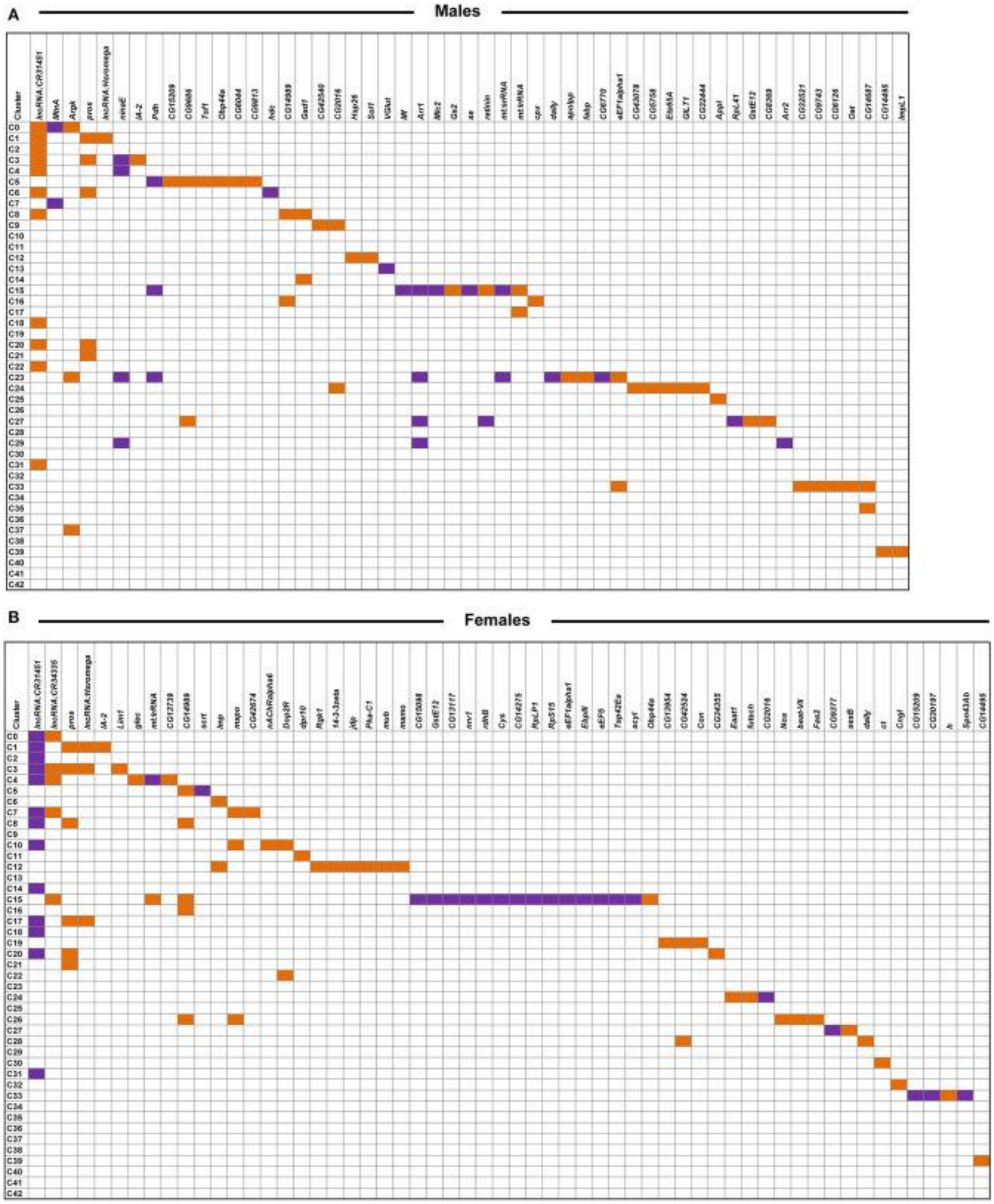

A
Males
Cluster
lncRNA:CR31451
MtnA
Argk
pros
lncRNA:Hsromega
ninaE
IA-2
Pdh
CG15309
CG9686
Tsf1
Obp44a
CG6044
CG9813
hdc
CG14989
Gad1
CG42540
CG2016
Hsp26
Sod1
VGlut
Mf
Arr1
Mlc2
Gs2
ae
retinin
mt:srRNA
mt:lrRNA
cpx
dally
apolpp
fabp
CG6770
eEF1alpha1
CG43078
CG5758
Elo68A
GILT1
CG32444
Appl
RpL41
GstE12
CG8369
Arr2
CG32521
CG9743
CG6126
Gat
CG14687
CG14485
lmpL1
C0
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
C17
C18
C19
C20
C21
C22
C23
C24
C25
C26
C27
C28
C29
C30
C31
C32
C33
C34
C35
C36
C37
C38
C39
C40
C41
C42
B
Females
Cluster
lncRNA:CR31451
lncRNA:CR34335
pros
lncRNA:Hsromega
IA-2
Lim1
glec
mt:lrRNA
CG17739
CG14989
scrt
Imp
mapo
CG42674
nAChRalpha6
Dop2R
dpr10
Rgk1
14-3-3zeta
Jdp
Pka-C1
mub
mamo
CG15098
GstE12
CG13117
nrv1
rdhB
Cys
CG14275
RpLP1
RpS15
eEF1alpha1
EbpII
eEF5
Tsp42Ea
scyl
Obp44a
CG13654
CG42524
Con
CG34355
Eaat1
futsch
CG2016
Nos
beat-VII
Fas3
CG9377
sesB
dally
ct
Cngl
CG15309
CG30197
h
Spn43Ab
CG14485
C0
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
C17
C18
C19
C20
C21
C22
C23
C24
C25
C26
C27
C28
C29
C30
C31
C32
C33
C34
C35
C36
C37
C38
C39
C40
C41
C42

**Figure 3.5. Differentially expressed genes across clusters in males (A) and females (B) after developmental alcohol exposure.** Differentially expressed genes are listed on the top (columns) and cell clusters are represented by the rows. Upregulated genes are indicated with orange and downregulated genes are indicated with purple. Differentially expressed genes are filtered at $|\log_e FC| > 0.25$ and a Bonferroni adjusted $p < 0.05$. Differentially expressed genes that survive a threshold of $|\log_e FC| > 1.0$ with a Bonferroni adjusted $p < 0.05$ are shown in Supplementary Figure S3.2.

We assessed global interaction networks of differentially expressed gene products across all cell clusters for males and females separately (Figure 3.6). The male interaction network is composed of modules associated with glutathione metabolism, lipid transport, glutamate and GABA metabolism, and vision (Figure 3.6A). The female interaction network also contains modules associated with glutamate and GABA metabolism, lipid metabolism, and vision, but the composition of these modules is distinct from their male counterparts. In addition, the female network features modules associated with monoaminergic signaling, cell adhesion, and Wnt signaling (Figure 3.6B). Multiple cell clusters contribute to each network module, indicating that modulation of gene regulation by developmental alcohol exposure is coordinated across different cells throughout the brain.

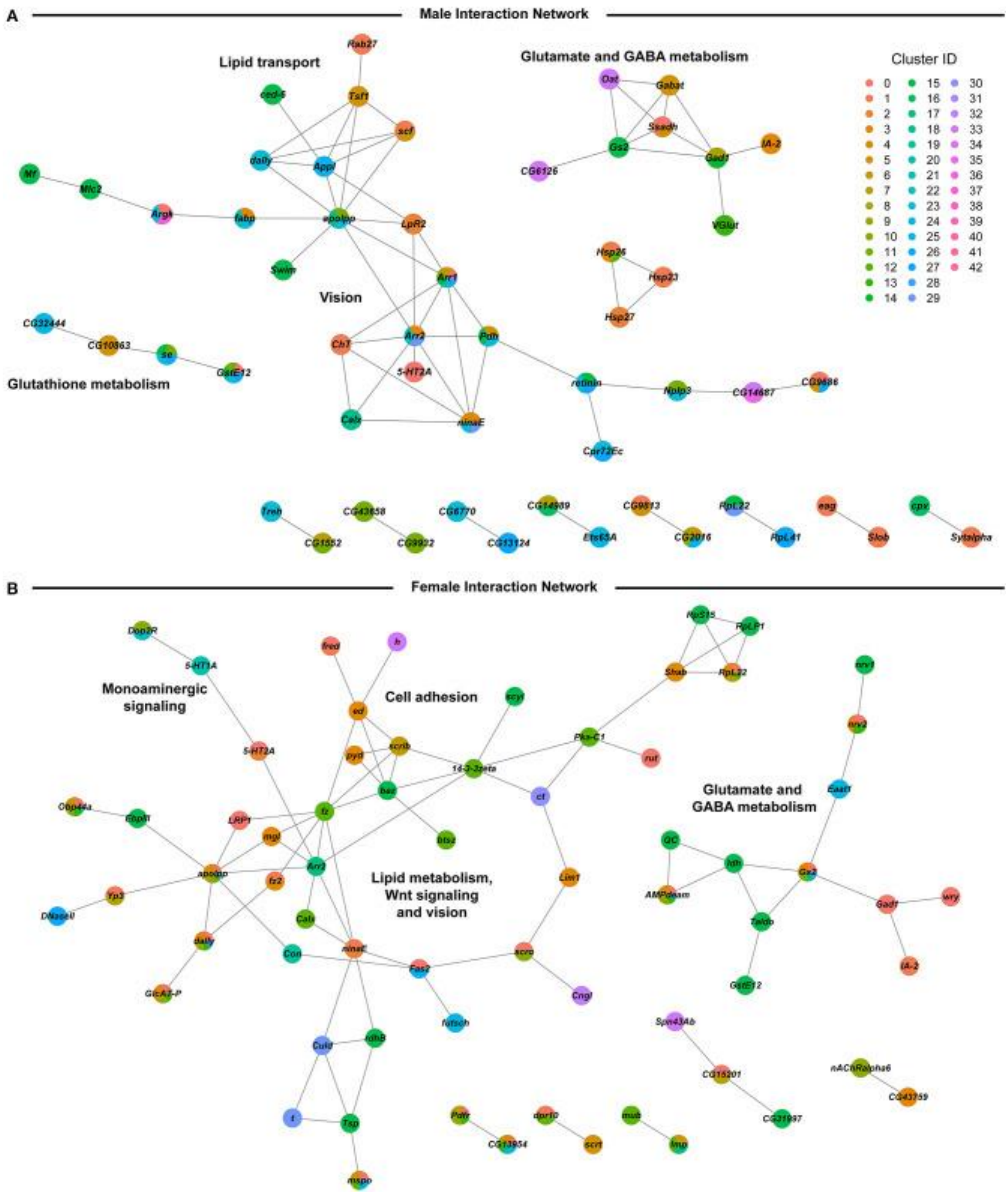

A
Male Interaction Network
Lipid transport
Glutamate and GABA metabolism
Vision
Glutathione metabolism
Rab27
Tsf1
ced-6
scf
daily
Appl
Mf
Mlc2
Argk
fabp
apolpp
LpR2
Swim
Arr1
Arr2
Cht7
Pdh
5-HT2A
Calx
ninaE
retinin
Nplp3
CG14687
CG9686
Cpr72Ec
CG32444
CG10863
se
GstE12
Oat
Gabat
Ssadh
Gs2
Gad1
IA-2
CG6126
VGlut
Hsp26
Hsp23
Hsp27
Treh
CG1552
CG43658
CG9922
CG6770
CG13124
CG14989
Ets65A
CG9813
CG2016
RpL22
RpL41
eag
Slob
cpx
Sytalpha
Cluster ID
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
30 31 32 33 34 35 36 37 38 39 40 41 42
B
Female Interaction Network
Monoaminergic signaling
Cell adhesion
Lipid metabolism, Wnt signaling and vision
Glutamate and GABA metabolism
Dop2R
5-HT1A
5-HT2A
fred
h
ed
scyl
pyd
scrib
14-3-3zeta
baz
Obp44a
Ebp III
LRP1
fz
mgl
btsz
apolpp
fz2
Arr2
Yp3
DNaseII
Calx
daily
Con
ninaE
Fas2
GlcAT-P
futsch
Cul4
rdhB
t
Tsp
mspo
RpS15
RpLP1
Shab
RpL22
Pka-C1
rut
ct
nrv1
nrv2
Eaat1
Lim1
scro
Cngl
GC
Idh
AMPdeam
Gs2
Gad1
wry
Taldo
GstE12
IA-2
Spn43Ab
CG15201
CG31997
nAChRalpha6
CG43759
Pdfr
CG13954
dpr10
scrt
mub
lmp

**Figure 3.6. Global interaction networks of differentially expressed gene products in males (A) and females (B) following developmental alcohol exposure.** Colors of the nodes correspond to the clusters in which expression of the gene is altered after growth on alcohol-supplemented medium.

We noticed that many genes that are differentially expressed following developmental exposure to ethanol correspond to genes that undergo altered expression when flies are exposed to cocaine (19). However, the transcriptional response to acute exposure to cocaine is larger than the transcriptional response to developmental alcohol exposure. Nonetheless, 69.7% of differentially expressed genes in males and 43.2% of differentially expressed genes in females in our data overlap with differentially expressed genes after consumption of cocaine (Figure 3.7 and Supplementary Table S3.7), although the magnitude and direction of differential expression of common genes between the two treatments varies by cell type (Supplementary Table S3.8). Gene ontology enrichment analyses of this common set of genes in each sex highlights gene ontology categories associated with development and function of the nervous system (Supplementary Table S3.9) (31).

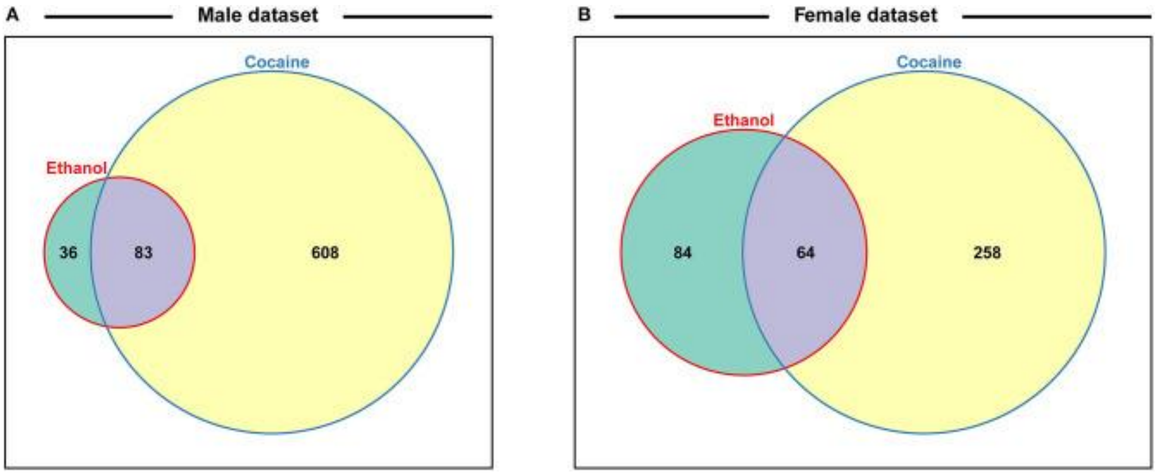


**Figure 3.7. Venn diagrams indicating the proportions of differentially regulated genes after exposure to alcohol during development or acute consumption of cocaine for males (A) and females (B).** Data for cocaine exposure are from ref 19. See also Supplementary Table S3.7.

## Discussion

We characterized the consequences of developmental alcohol exposure in Drosophila on viability, behavioral phenotypes, and gene expression in the brain. Characteristic features of FASD in humans include craniofacial dysmorphologies and cognitive impairments. Although we did not perform detailed morphometric measurements, we did not observe any overt morphological aberrations, and cognitive impairments are challenging to assess in Drosophila. Nevertheless, flies exposed to alcohol during embryonic and larval development showed changes in activity and sleep patterns (Figures 3.2C–G), consistent with previously observed effects of larval ethanol exposure on adult circadian rhythms (32, 33) and reminiscent of activity and sleep disturbances seen in children with FASD

(27, 28). We also find that growth on alcohol supplemented medium results in reduced ethanol sensitivity of adult flies, in agreement with a previous study (Figure 3.2B) (14).

We hypothesize that the effects of developmental alcohol exposure on changes in gene expression in the Drosophila central nervous system will converge on evolutionarily conserved cellular processes. Drosophila is advantageous for studies on gene expression at single cell resolution because we can survey the entire brain in a single analysis, unlike studies in rodents, and pooling multiple brains of the same genotype averages individual variation. The power to detect changes in gene expression in our study is improved by only considering changes in gene expression that are consistent across replicates.

We observed changes in gene expression in adult flies, even though exposure to alcohol occurred only during the larval stages and briefly after eclosion, after which adults were collected and maintained on regular medium without alcohol. It is possible that developmental alcohol exposure may result in epigenetic modifications that give rise to altered gene expression patterns into adulthood (34).

We observe changes in gene expression in diverse neuronal and glial cell populations (Figure 3.5). Since we are not able to sample all cells of the brain, it is likely that some neuronal or glial cell populations are not represented in our data. However, the major regions of the Drosophila brain and all major neurotransmitter cell types are represented (Figure 3.4 and Supplementary Table S3.3). The effects of developmental alcohol exposure are sexually dimorphic, similar to previously observed changes in transcript abundances following consumption of cocaine (19). Sexual dimorphism is also a hallmark of FASD, with different effects of fetal alcohol exposure on neural development and cognitive abilities between males and females (35–38). Although different genes are affected in males and females, gene ontology analysis indicates that they converge on the same biological processes, related to development and function of the nervous system (Supplementary Table S3.8). The considerable overlap between differentially expressed genes in response to alcohol and cocaine suggests common neural substrates that respond to toxic exposures. Genes associated with immune defense and xenobiotic detoxification, including the glutathione pathway, feature in interaction networks of differentially expressed gene products (Figure 3.6).

*lncRNA:CR31451* shows large sexually antagonistic responses to developmental alcohol exposure in many neuronal cell populations. Whereas, a previous study documented expression of this gene in glia (39), we only observe differential gene expression of *lncRNA:CR31451* in neurons under the conditions of our study (Figure 3.5). Future studies are needed to assess whether this gene product fulfills a regulatory function that affects multiple neurotransmitter signaling processes and whether its sex-antagonistic response to alcohol exposure could in part cause the differential gene expression patterns seen in males and females.

Our observations of extensive changes in gene expression in glia in response to developmental alcohol exposure are in accordance with the role of glia in FASD. Fetal alcohol exposure leads to impaired astrocyte development and differentiation, which gives rise to microencephaly (40, 41). In addition, ethanol exposure increases permeability of the blood brain barrier (42), which in Drosophila is formed by the surface glia (43). Among the glial genes that show altered expression after developmental alcohol exposure in Drosophila are *GILT1*, which contributes to the immune defense response to bacteria (44), *Gs2* and *Eaat1*, which are involved in glutamine synthesis and transport of glutamate in astrocytes (45, 46), *GstE12* and *se*, which are involved in glutathione metabolism (47), and *fabp* and *apolpp*, which function in lipid metabolism (48, 49). GABA signaling and glutamate signaling neuronal cell populations feature prominently in our data (Figure 4). Glutamate is also a precursor for the biosynthesis of glutathione, which is produced in glia and protects against oxidative stress and detoxification of xenobiotics (50). Developmental alcohol exposure interferes with glutamate and GABA signaling because ethanol is both an antagonist to the NMDA glutamate receptor and mimics GABA (51). Consequently, fetal alcohol exposure results in neuronal apoptosis during the rapid brain growth spurt during which the astrocytes play a major role (51, 52). Evolutionarily conserved neural processes that respond to developmental alcohol exposure in Drosophila thus provide a blueprint for translational studies on alcohol-induced effects on gene expression in the brain that may contribute to or result from FASD in human populations.

## Acknowledgements

This work was supported by grants GM128974 and DA041613 from the National Institutes of Health to TFCM and RRHA. The funding sources were not involved in the study design; collection, analysis, or interpretation of data; writing of the report, or decision to submit the article for publication.

## References

1. Jones K, Smith D. Recognition of the fetal alcohol syndrome in early infancy. *The Lancet* (1973) 302: 999-1001. doi: 10.1016/s0140-6736(73)91092-1.

2. Hoyme HE, May PA, Kalberg WO, Kodituwakku P, Gossage JP, Trujillo PM, et al. A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: clarification of the 1996 Institute of Medicine criteria. *Pediatrics* (2005) 115: 39-47. doi: 10.1542/peds.2004-0259.

3. Roozen S, Peters GJ, Kok G, Townend D, Nijhuis J, Curfs L. Worldwide prevalence of Fetal Alcohol Spectrum Disorders: A systematic literature review including meta-analysis. *Alcohol Clin Exp Res*. (2016) 40:18-32. doi: 10.1111/acer.12939.

4. Kaminen-Ahola N. Fetal alcohol spectrum disorders: genetic and epigenetic mechanisms. *Prenat Diagn*. (2020) 40: 1185-92. doi: 10.1002/pd.5731.

5. Clarren SK, Smith DW. The fetal alcohol syndrome. *Lamp.* (1978) 35: 4-7.

6. Pulsifer MB. The neuropsychology of mental retardation. *J Int Neuropsychol Soc.* (1996) 2: 159-76. doi: 10.1017/s1355617700001016.

7. Eckardt MJ, File SE, Gessa GL, Grant KA, Guerri C, Hoffman PL, et al. Effects of moderate alcohol consumption on the central nervous system. *Alcohol Clin Exp Res.* (1998) 22: 998-1040. doi: 10.1111/j.1530-0277.1998.tb03695.x.

8. Lange S, Rehm J, Anagnostou E, Popova S. Prevalence of externalizing disorders and Autism Spectrum Disorders among children with Fetal Alcohol Spectrum Disorder: systematic review and meta-analysis. *Biochem Cell Biol.* (2018) 96: 241-51. doi: 10.1139/bcb-2017-0014.

9. Tan CH, Denny CH, Cheal, NE, Sniezek JE, Kanny D. Alcohol use and binge drinking among women of childbearing age - United States, 2011-2013. *MMWR*. (2015) 64: 1042-6. doi: 10.15585/mmwr.mm6437a3.

10. Morozova TV, Goldman D, Mackay TFC, Anholt RRH. The genetic basis of alcoholism: multiple phenotypes, many genes, complex networks. *Genome Biol*. (2012) 13: 239. doi: 10.1186/gb-2012-13-2-239.

11. Morozova TV, Mackay TFC, Anholt RRH. Genetics and genomics of alcohol sensitivity. *Mol Genet Genomics* (2014) 289: 253-69. doi: 10.1007/s00438-013-0808-y.

12. Scholz H. Unraveling the mechanisms of behaviors associated with AUDs using flies and worms. *Alcohol Clin Exp Res.* (2019) 43: 2274-84. doi: 10.1111/acer.14199.

13. Petruccelli E, Kaun KR. Insights from intoxicated Drosophila. *Alcohol* (2019) 74: 21-7. doi: 10.1016/j.alcohol.2018.03.004.

14. McClure KD, French RL, Heberlein U. A Drosophila model for fetal alcohol syndrome disorders: role for the insulin pathway. *Dis Model Mech.* (2011) 4: 335-46. doi: 10.1242/dmm.006411.

15. Morozova TV, Hussain Y, McCoy LJ, Zhirnov EV, Davis MR, Pray VA, et al. A *Cyclin E* centered genetic network contributes to alcohol-induced variation in Drosophila development. *G3 (Bethesda)* (2018) 8: 2643-53. doi: 10.1534/g3.118.200260.

16. Scepanovic G, Stewart BA. Analysis of Drosophila nervous system development following an early, brief exposure to ethanol. *Dev Neurobiol.* (2019) 79: 780-93. doi: 10.1002/dneu.22718.

17. Logan-Garbisch T, Bortolazzo A, Luu P, Ford A, Do D, Khodabakhshi P, et al. Developmental ethanol exposure leads to dysregulation of lipid metabolism and oxidative stress in Drosophila. *G3 (Bethesda)* (2014) 5: 49-59. doi: 10.1534/g3.114.015040.

18. Guevara A, Gates H, Urbina B, French R. Developmental ethanol exposure causes reduced feeding and reveals a critical role for Neuropeptide F in survival. *Front Physiol.* (2018) 9: 237. doi: 10.3389/fphys.2018.00237.

19. Baker BM, Mokashi SS, Shankar V, Hatfield JS, Hannah RC, Mackay TFC, et al. The Drosophila brain on cocaine at single cell resolution. *Genome Res.* (2021) Epub Ahead of print. doi: 10.1101/gr.268037.120.

20. Sass TN, MacPherson RA, Mackay TFC, Anholt RRH. A high-throughput method for measuring alcohol sedation time of individual *Drosophila melanogaster. J Vis Exp.* (2020) 158: e61108. doi: 10.3791/61108.

21. Cichewicz K, Hirsh J. ShinyR-DAM: A program analyzing Drosophila activity, sleep and circadian rhythms. *Commun Biol.* (2018) 1:1-5. doi:10.1038/s42003-018-0031-9.

22. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* (2018) 36: 411-20. doi: 10.1038/nbt.4096.

23. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* (2019) 20: 296. doi: 10.1186/s13059-019-1874-1.

24. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* (2015)16: 278. doi: 10.1186/s13059-015-0844-5.

25. Doncheva NT, Morris J, Gorodkin J, Jensen LJ. Cytoscape stringApp: Network analysis and visualization of proteomics data. *J Proteome Res.* (2018) 18: 623-32. doi: 10.1021/acs.jproteome.8b00702.

26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13: 2498-504. doi: 10.1101/gr.1239303.

27. Hanlon-Dearman A, Chen ML, Olson HC. Understanding and managing sleep disruption in children with fetal alcohol spectrum disorder. *Biochem Cell Biol.* (2018) 96: 267-74. doi: 10.1139/bcb-2017-0064.

28. Kamara D, Beauchaine TP. A review of sleep disturbances among infants and children with neurodevelopmental disorders. *Rev J Autism Dev Disord.* (2020) 7: 278-94. doi: 10.1007/s40489-019-00193-8.

29. Becht E, McInnes L, Healy J, Dutertre C, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* (2019) 37: 38-44. doi:10.1038/nbt.4314

30. Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, et al.. FlyBase 2.0: the next generation. *Nucleic Acids Res.* (2019) 47:D759–65. 10.1093/nar/gky1003

31. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* (2003) 13: 2129-41. doi: 10.1101/gr.772403.

32. Seggio JA, Possidente B, Ahmad ST. Larval ethanol exposure alters adult circadian free-running locomotor activity rhythm in *Drosophila melanogaster*. *Chronobiol Int.* (2012) 29:75–81. 10.3109/07420528.2011.635236

33. Ahmad ST, Steinmetz SB, Bussey HM, Possidente B, Seggio JA. Larval ethanol exposure alters free-running circadian rhythm and *per* locus transcription in adult *D. melanogaster period* mutants. *Behav Brain Res.* (2013) 241:50–5. 10.1016/j.bbr.2012.11.035

34. Kobor MS, Weinberg J. Focus on: epigenetics and fetal alcohol spectrum disorders. *Alcohol Res Health* (2011) 34: 29-37.

35. May PA, Tabachnick B, Hasken JM, Marais AS, de Vries MM, Barnard R, et al. Who is most affected by prenatal alcohol exposure: Boys or girls? *Drug Alcohol Depend.* (2017) 177: 258-67. doi: 10.1016/j.drugalcdep.2017.04.010.

36. Coleman LG Jr, Oguz I, Lee J, Styner M, Crews FT. Postnatal day 7 ethanol treatment causes persistent reductions in adult mouse brain volume and cortical neurons with sex specific effects on neurogenesis. *Alcohol* (2012) 46: 603-12. doi: 10.1016/j.alcohol.2012.01.003.

37. Weinberg J, Sliwowska JH, Lan N, Hellemans KG. Prenatal alcohol exposure: foetal programming, the hypothalamic-pituitary-adrenal axis and sex differences in outcome. *J Neuroendocrinol.* (2008) 20:470-88. doi: 10.1111/j.1365-2826.2008.01669.x.

38. Kelly SJ, Leggett DC, Cronise K. Sexually dimorphic effects of alcohol exposure during development on the processing of social cues. *Alcohol Alcohol.* (2009) 44: 555-60. doi: 10.1093/alcalc/agp061.

39. Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft Ł, et al. A single-cell transcriptome atlas of the aging Drosophila brain. *Cell* (2018) 174: 982-98.e20. doi:10.1016/j.cell.2018.05.057

40. Wilhelm CJ, Guizzetti M. Fetal Alcohol Spectrum Disorders: An overview from the glia perspective. *Front Integr Neurosci.* (2016) doi: 10.3389/fnint.2015.00065.

41. Guerri C, Pascual M, Renau-Piqueras J. Glia and fetal alcohol syndrome. *Neurotoxicol.* (2001) 22: 593-9. doi: 10.1016/s0161-813x(01)00037-7

42. Tufa U, Ringuette D, Wang XF, Levi O, Carlen P. Two-photon microscopy imaging of blood brain barrier leakage in fetal alcohol disorder mice. *Optics in the Life Sciences Congress, OSA Technical Digest (online) (Optical Society of America)* (2017) paper JTu4A.9.

43. DeSalvo MK, Hindle SJ, Rusan ZM, Orng S, Eddison M, Halliwill K, et al. The Drosophila surface glia transcriptome: evolutionary conserved blood-brain barrier processes. *Front Neurosci.* (2014) 8: 346. doi: 10.3389/fnins.2014.00346.

44. Kongton K, McCall K, Phongdara A. Identification of gamma-interferon-inducible lysosomal thiol reductase (GILT) homologues in the fruit fly *Drosophila melanogaster*. *Dev Comp Immunol.* (2014) 44: 389-96. doi: 10.1016/j.dci.2014.01.007.

45. Parinejad N, Peco E, Ferreira T, Stacey SM, van Meyel DJ. Disruption of an EAAT-mediated chloride channel in a Drosophila model of ataxia. *J Neurosci.* (2016) 36: 7640-7. doi:10.1523/JNEUROSCI.0197-16.2016

46. Chen WT, Yang HY, Lin CY, Lee YZ, Ma SC, Chen WC, et al. Structural insight into the contributions of the N-terminus and key active-site residues to the catalytic efficiency of glutamine synthetase 2. *Biomolecules* (2020) 10:1671. doi:10.3390/biom10121671

47. Saisawang C, Wongsantichon J, Ketterman AJ. A preliminary characterization of the cytosolic glutathione transferase proteome from *Drosophila melanogaster*. *Biochem J.* (2012) 442:181-90. doi:10.1042/BJ20111747

48. Palm W, Sampaio JL, Brankatschk M, Carvalho M, Mahmoud A, Shevchenko A, et al, Lipoproteins in *Drosophila melanogaster* -- assembly, function, and influence on tissue lipid composition. *PLoS Genet.* (2012) 8: e1002828. doi:10.1371/journal.pgen.1002828

49. Gerstner JR, Vanderheyden WM, Shaw PJ, Landry CF, Yin JC. Fatty-acid binding proteins modulate sleep and enhance long-term memory consolidation in Drosophila. *PLoS One* (2011) 6: e15890. doi:10.1371/journal.pone.0015890

50. Sedlak TW, Paul BD, Parker GM, Hester LD, Snowman AM, Taniguchi Y, et al. The glutathione cycle shapes synaptic glutamate activity. *Proc. Natl. Acad. Sci. USA.* (2019) 116: 2701-6. doi:10.1073/pnas.1817885116

51. Olney JW, Wozniak DF, Jevtovic-Todorovic V, Farber NB, Bittigau P, Ikonomidou C. Glutamate and GABA receptor dysfunction in the fetal alcohol syndrome. *Neurotox Res.* (2002) 4: 315-25. doi:10.1080/1029842021000010875

52. Tsai G, Gastfriend DR, Coyle JT. The glutamatergic basis of human alcoholism. *Am J Psychiatry* (1995) 152: 332-40. Ddoi:10.1176/ajp.152.3.332

# CHAPTER FOUR

# MODULATION OF THE DROSOPHILA TRANSCRIPTOME BY DEVELOPMENTAL EXPOSURE TO ALCOHOL

Morozova TV*, Shankar V*, MacPherson RA, Mackay TFC, Anholt RRH. Modulation of the Drosophila transcriptome by developmental exposure to alcohol. *BMC Genomics*. 2022 May 6;23(1):347. doi: 10.1186/s12864-022-08559-9. PMID: 35524193; PMCID: PMC9074282.

*contributed equally

## Author Contribution Statement

My contributions to the work described in this chapter are as follows: I was primarily involved in the behavioral studies of this chapter, where I reared and collected the Drosophila, performed the behavioral studies, and analyzed the resulting behavioral data.

## Introduction

Prenatal exposure to ethanol can trigger a wide range of adverse physiological, behavioral, and cognitive outcomes, referred to as fetal alcohol spectrum disorder (FASD) [1–4]. Fetal alcohol syndrome (FAS) is the most severe FASD. Affected individuals show craniofacial defects, deficiencies in cognition and behavioral anomalies, including hyperactivity, attention deficit disorder and motor coordination [1, 5–7].

FAS/FASD is the most common preventable pediatric disorder, often diagnostically confounded with autism spectrum disorder [8]. The Centers for Disease Control and Prevention found that, despite warning labels on alcoholic beverages that indicate possible adverse effects on prenatal development, 1 in 10 pregnant women report alcohol use and more than 3 million women in the USA are at risk of exposing their developing fetus to alcohol [9]. Although defects from prenatal alcohol exposure can be replicated in mouse models [10], identifying genetic factors that contribute to susceptibility to FASD is virtually impossible in human populations since time, dose, and frequency of exposure are generally unknown, and manifestations of FASD are diverse and become evident long after exposure.

*Drosophila melanogaster* presents an advantageous model for studies on the genetic underpinnings associated with symptoms of developmental alcohol exposure. The Drosophila model allows strict control over the genetic background. In addition, individuals of the same genotype can be reared in large numbers under controlled environmental conditions, without regulatory restrictions and at low cost. Following acute exposure to alcohol, flies undergo loss of postural control, sedation, and development of tolerance, physiological and behavioral changes that resemble human alcohol intoxication [11–14].

Previous studies on the effects of developmental alcohol exposure in Drosophila showed reduced viability and delayed development time [15, 16], reduced adult body size [15], and disruption of neural development [17]. Developmental exposure to alcohol was associated with reduction in the expression of a subset of insulin-like peptides and the insulin receptor [15], dysregulation of lipid metabolism and concomitant increased oxidative stress [18] and reduced larval food intake due to altered neuropeptide F signaling [19].

In previous studies, we have taken advantage of natural variation in the *Drosophila melanogaster* Genetic Reference Panel (DGRP), a well characterized population of 205 inbred wild-derived lines with complete genome sequences [20, 21], to study the genetic underpinnings of developmental alcohol exposure [16], voluntary ethanol consumption [22], acute ethanol intoxication, and induction of tolerance [23, 24], a prelude to the development of alcohol dependence in people. Linkage disequilibrium in the DGRP decays within a few hundred base pairs [21], which enables identification of candidate causal single nucleotide polymorphisms (SNPs) associated with phenotypic variation in alcohol-related traits. A major unresolved question relates to the mechanism(s) by which alcohol exposure during development affects adult phenotypes. Here, we performed RNA sequencing to assess the effects of developmental alcohol exposure on genome wide genetic variation in gene expression of young adults. Evolutionary conservation of fundamental biological processes and superposition of human orthologs on transcriptional networks identified in flies provides translational potential for studies in the Drosophila model.

## Results

### *Transcriptional profiles of flies reared on ethanol supplemented medium*

We performed transcriptional profiling of flies reared on regular medium or medium supplemented with 10% (v/v) ethanol from 96 DGRP lines that span the phenotypic spectrum of alcohol sensitivity following developmental exposure to alcohol (Supplementary File S4.1) [16]. For each line, we obtained duplicate samples of 30 males and 25 females, aged 3–5 days, all collected at the same time of day, and performed RNA sequencing using 125 bp single end reads. A total of 33,580 transcripts were expressed in adult flies, of which 16,165 (48.1%) are annotated and 17,415 (51.9%) are novel transcribed regions (NTRs) [25, 26] (Supplementary File S4.2). We performed three-way factorial mixed model ANOVAs with the main effects of DGRP line, sex, and treatment for all expressed transcripts, and used a False Discovery Rate (FDR) threshold $< 0.05$ for statistical significance of each term in the ANOVA model (Supplementary File S4.2). As in previous analyses of genome-wide gene expression using whole flies [25–27], we find that expression of nearly all expressed transcripts is sexually dimorphic (28,343, 84.4%). The expression of 16,278 transcripts (48.5%) is modulated by alcohol, and for 10,002 transcripts the transcriptional response to ethanol differs between males and females (*i.e.*, there is a treatment by sex interaction). There is significant genetic variation in expression for 10,620 transcripts, as well as context-specific genetic variation, with 11,338 transcripts exhibiting genetic variation in sexual dimorphism (sex by line interaction), 1,222 showing genetic variation in response to developmental exposure to ethanol (treatment by line interaction) and 77 with genetic variation in sexual dimorphism in response to ethanol (sex by treatment by line interaction).

Because we found extensive interactions with sex and treatment and sex and DGRP line, we performed two-way ANOVAs partitioning gene expression variation by line, treatment, and the line by treatment interaction ($L \times T$), separately for males and females (Supplementary File S4.3). The main effect of treatment was significant ($\mathrm{FDR} < 0.05$) for 14,158 (13,827) transcripts in females (males), and the main effect of line was significant for 13,521 (20,996) transcripts in females (males). We are most interested in transcripts with a significant genotype by alcohol treatment interaction ($L \times T$) since these transcripts show genetic variation in their response to developmental ethanol exposure. When we compared expression profiles of flies grown on ethanol supplemented medium with transcript abundance levels obtained under standard growth conditions, we found 939 significant $L \times T$ interactions in females and 823 in males (Supplementary File S4.4). Of these 1,351 transcripts that have genetically variable responses to ethanol exposure during development, 253 are in common between males and females, 499 are female-specific and 346 are male-specific. The transcripts with significant $L \times T$ terms in females were enriched for biological process gene ontology terms involved in metabolism, biosynthesis, transcription, immune/defense response, and chromatin organization; and the transcripts with significant $L \times T$ terms in males were enriched for biological process gene ontology terms involved in immune/defense response, metabolism, and development [28] ( Supplementary File S4.4). A total of 65.6% of these genes have human orthologs ( Supplementary File S4.4).

### *Co-regulated modules of transcripts that are differentially regulated after developmental alcohol exposure*

We analyzed the sex-specific correlation structure of the developmental alcohol-sensitive transcriptome and identified eight highly interconnected modules in females (Figure 4.1). These modules contained transcripts associated with xenobiotic detoxification and metabolism (Figures 4.1A, B), development and cell adhesion (Figures 4.1B, C), cuticle formation (Figure 4.1E) and neural signaling (Figure 4.1G). The latter includes *sNPF*, which is associated with neuropeptide F signaling and has been previously implicated with reduced larval food intake when larvae were grown on alcohol-supplemented medium [19]. One module consists almost exclusively of NTRs (designated XLOCs; Figure 4.2D) [25, 26]. Of special interest is *Ilp3*, which encodes an insulin-like peptide implicated in developmental exposure to alcohol [15] and is correlated with a network of NTRs (Figure 4.1F). The most highly correlated group of transcripts are the small nucleolar RNAs (snoRNAs; Figure 4.1H).

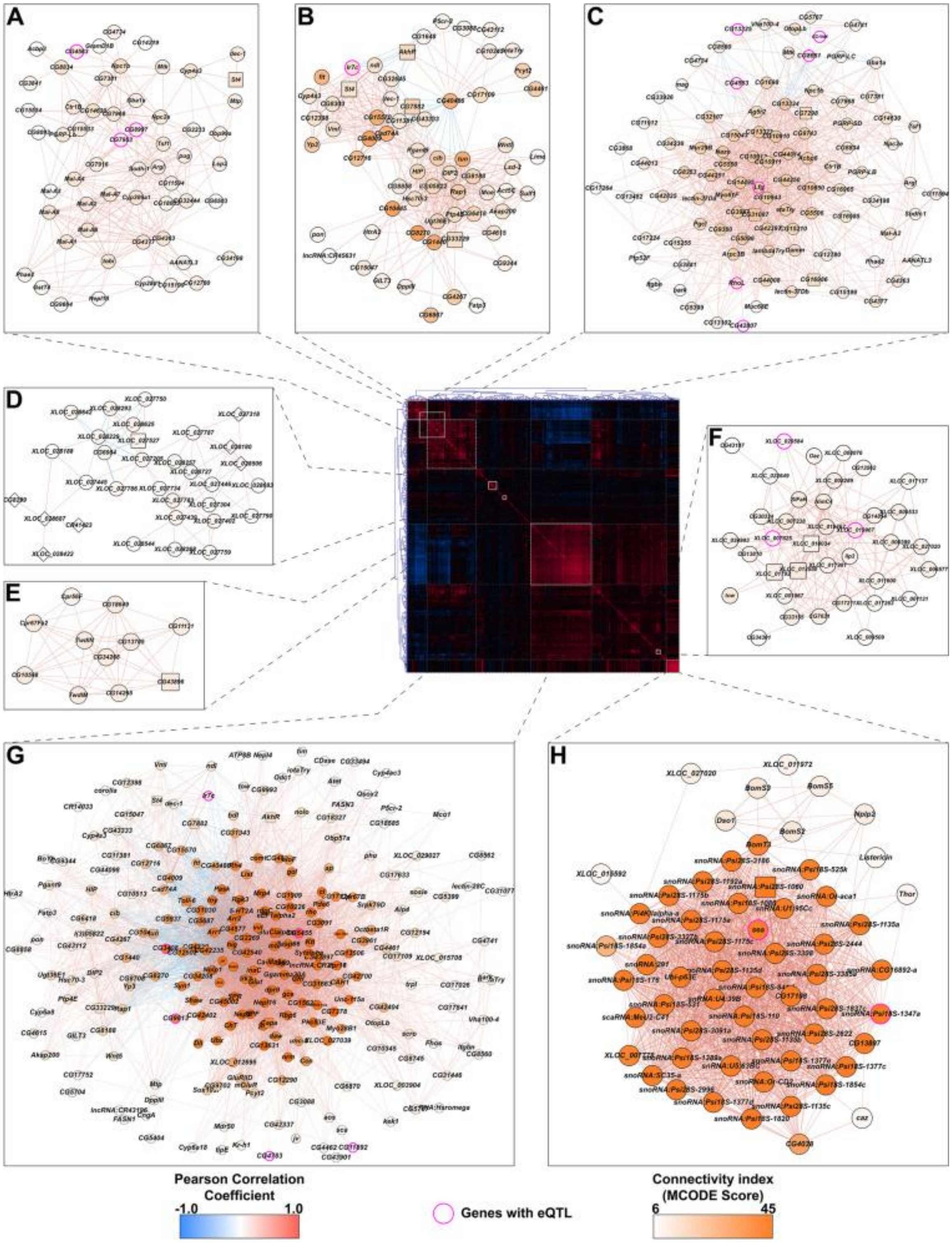

A
B
C
D
E
F
G
H
Pearson Correlation Coefficient
-1.0
1.0
Genes with eQTL
Connectivity index (MCODE Score)
6
45

**Figure 4.1. Correlations of differences in gene expression between developmental ethanol treatment and control in females.** The center panel heat map corresponds to the unfiltered, bi-clustered correlation matrix calculated for differences in expression of genes with a statistically significant line-by-ethanol treatment interaction term in a linear mixed effects model. The strength of the correlation is depicted as gradients and the directionality as color (positive correlations in red and negative correlations in blue). Networks derived from clusters with strong intra-connectivity are depicted around the center panel (panels **A-H**). The MCODE connectivity score for each node is represented as a color gradient. The edge colors follow the same scheme as the center panel (strength as gradient and directionality as color). Genes with statistically significant eQTLs are highlighted with pink borders.

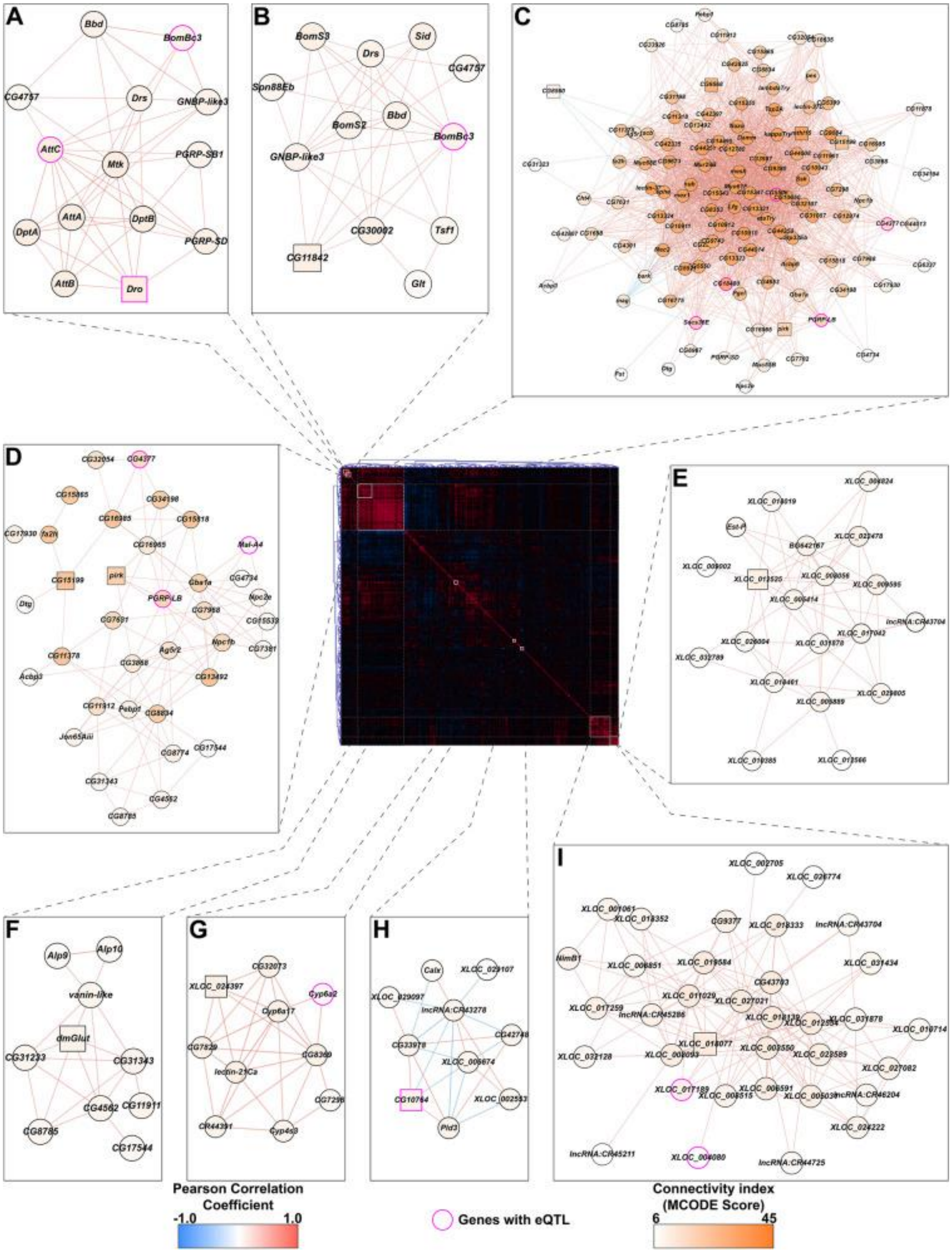

A
Bbd
BomBc3
CG4757
Drs
GNBP-like3
AttC
Mtk
PGRP-SB1
AttA
DptB
DptA
PGRP-SD
AttB
Dro
B
BomS3
Sid
Drs
CG4757
Spn88Eb
Bbd
BomS2
BomBc3
GNBP-like3
CG30002
Tsf1
CG11842
Glt
C
D
CG32054
CG4377
CG15065
CG34198
CG16985
CG15818
CG17930
fa2h
CG16965
Mal-A4
CG15199
pirk
Gba1a
CG4734
Dtg
PGRP-LB
Npc2e
CG7631
CG7958
CG15533
Npc1b
CG7381
CG11378
Ag5r2
CG3868
CG13492
Acbp3
CG11912
Pebp1
CG8834
Jon65Aiii
CG8774
CG17544
CG31343
CG4562
CG8785
E
XLOC_004824
XLOC_018019
Est-P
XLOC_022478
BG642167
XLOC_009002
XLOC_012525
XLOC_008056
XLOC_009595
XLOC_006414
lncRNA:CR43704
XLOC_026004
XLOC_031878
XLOC_017042
XLOC_032789
XLOC_016461
XLOC_029605
XLOC_008889
XLOC_019385
XLOC_013566
F
Alp9
Alp10
vanin-like
dmGlut
CG31233
CG31343
CG4562
CG11911
CG8785
CG17544
G
CG32073
XLOC_024397
Cyp6a2
Cyp6a17
CG7829
CG8369
lectin-21Ca
CG7298
CR44391
Cyp4s3
H
Calx
XLOC_029107
XLOC_029097
lncRNA:CR43278
CG33978
CG42748
XLOC_006674
CG10764
XLOC_002553
Pld3
I
XLOC_002705
XLOC_026774
XLOC_001061
XLOC_018352
CG9377
XLOC_018333
lncRNA:CR43704
NimB1
XLOC_006851
XLOC_019584
XLOC_031434
CG43703
XLOC_017259
XLOC_011029
XLOC_027021
lncRNA:CR45286
XLOC_018139
XLOC_012554
XLOC_031878
XLOC_019714
XLOC_032128
XLOC_018077
XLOC_003550
XLOC_008093
XLOC_028589
XLOC_027082
XLOC_017189
XLOC_006591
XLOC_005039
lncRNA:CR46204
XLOC_008515
XLOC_024222
lncRNA:CR45211
XLOC_004080
lncRNA:CR44725
Pearson Correlation Coefficient
-1.0
1.0
Genes with eQTL
Connectivity index (MCODE Score)
6
45

**Figure 4.2. Correlations of differences in gene expression between developmental ethanol treatment and control in males.** The center panel heat map corresponds to the unfiltered, bi-clustered correlation matrix calculated for differences in expression of genes with a statistically significant line-by-ethanol treatment interaction term in a linear mixed effects model. The strength of the correlation is depicted as gradients and the directionality as color (positive correlations in red and negative correlations in blue). Networks derived from clusters with strong intra-connectivity are depicted around the center panel (panels **A-I**). The MCODE connectivity score for each node is represented as a color gradient. The edge colors follow the same scheme as the center panel (strength as gradient and directionality as color). Genes with statistically significant eQTLs are highlighted with pink borders.

Although there was overlap between alcohol-modulated transcripts in males and females, the correlation structure for differentially expressed genes in males shows an entirely different modular organization (Figure 4.2). The nine male modules are generally smaller and more difficult to interpret than the female modules, as they are dominated by NTRs and genes of unknown function. Whereas the functions of this new class of non-coding transcripts remain to be established, they feature prominently in networks of alcohol-modulated transcripts (Figures 4.1 and 4.2). This suggests a regulatory role for non-coding elements in the genome in modulating the transcriptional response to developmental alcohol exposure.

To identify genetic associations within modules of differentially expressed genes, we mapped expression quantitative trait loci associated with the difference in expression between standard and ethanol-supplemented medium, or response expression-Quantitative Trait Loci (e-QTLs; [25, 26]). We identified 53 eQTLs, including 19 NTRs, in females, and 45 eQTLs, including 11 NTRs and four long-noncoding RNAs (lncRNAs), in males (Supplementary File S4.5). All eQTLs, with the exception of the eQTL associated with the difference in *NetA* expression between standard and ethanol-supplemented medium in females, were in *trans* (greater than 1 kb from the start and end of the gene body) to the genes associated with an *L*x*T* interaction). Therefore, we determined to what genes these eQTLs were in *cis* (defined by within 1 kb of a gene body) (Supplementary File S4.5). We then input the genes with *L*x*T* interactions and the genes to which the eQTLs are *cis* into known genetic and protein–protein or RNA–protein physical interaction networks to construct sex-specific networks associated with genetic variation in response to developmental exposure to ethanol (Figure 4.3). The resulting networks are composed of integrative modules that highlight considerable overlap between cellular processes in males and females despite significant sexual dimorphism in ethanol-induced differential gene expression. Genes associated with neuronal development and differentiation, response to ethanol, and reproduction are evident in networks of both sexes. The female network also incorporates modules associated with glutathione metabolism and phototransduction (Figure 4.3A), whereas the male network contains modules associated with immune response, starvation and stress response, and septate junction assembly (Figure 4.3B).

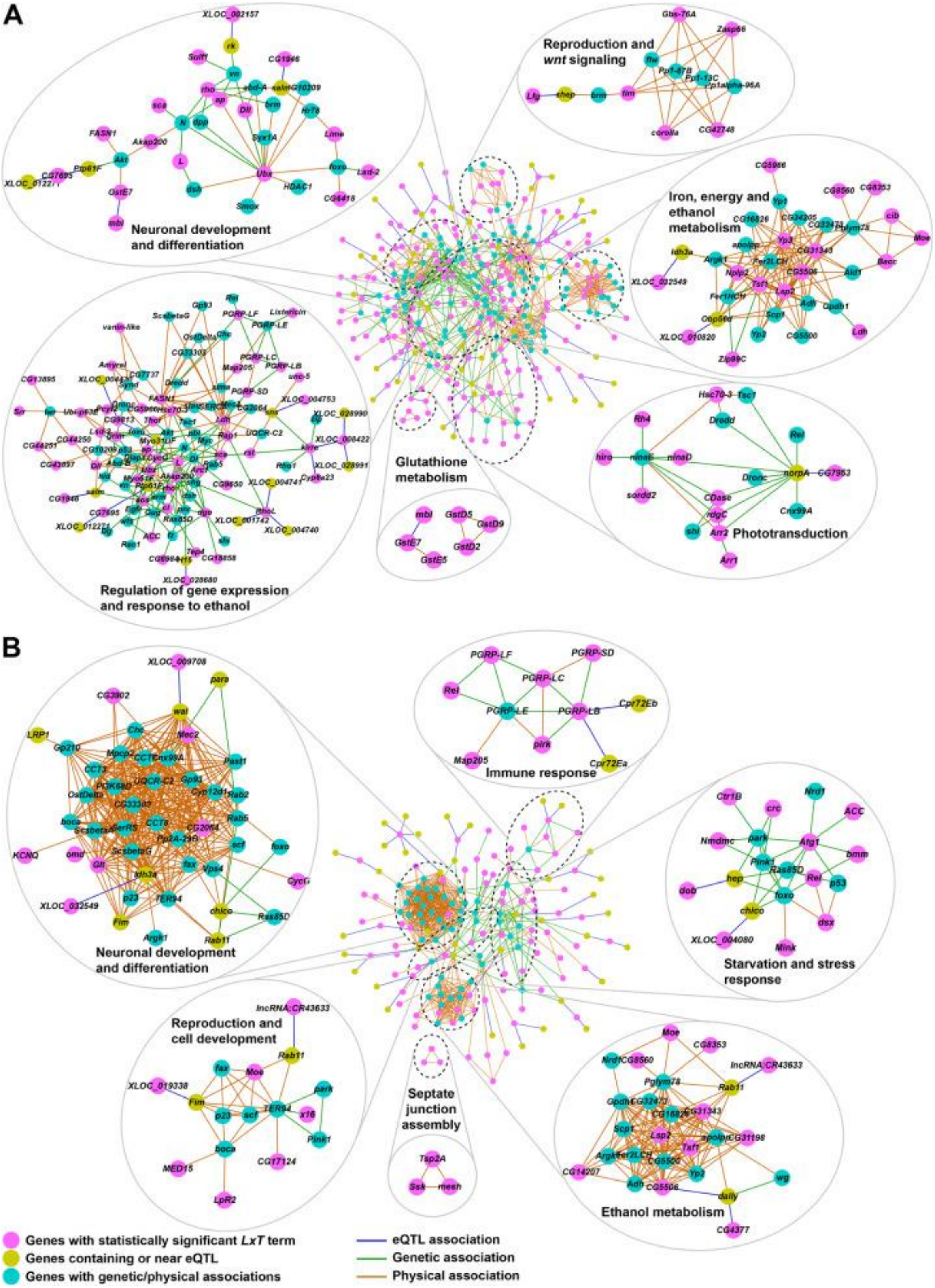

A
Neuronal development and differentiation
Reproduction and *wnt* signaling
Iron, energy and ethanol metabolism
Regulation of gene expression and response to ethanol
Glutathione metabolism
Phototransduction
B
Neuronal development and differentiation
Immune response
Starvation and stress response
Reproduction and cell development
Septate junction assembly
Ethanol metabolism
Genes with statistically significant *LxT* term
Genes containing or near eQTL
Genes with genetic/physical associations
eQTL association
Genetic association
Physical association

**Figure 4.3. Female (A) and male (B) interaction networks built from eQTLs and known genetic and physical associations.** The central networks in each panel represent the sex-separated, filtered interaction networks generated by incorporating eQTL associations calculated from expression differences, between ethanol and control conditions, of genes with a statistically significant line-by-treatment (*LxT*) term in the ANOVA model, to known genetic and protein–protein or RNA–protein physical interactions from the FlyBase interaction database. Pink nodes represent the genes from the *LxT* set. Yellow nodes represent genes either containing or within 1,000 bp of the eQTL variant. Cyan nodes represent genes with known genetic or physical interactions to the rest of the network. Blue edges represent the eQTL associations from this study. Green and orange edges represent known genetic and physical associations from the Flybase interaction database. Individual inlets of genes around the central network are MCODE-generated modules of genes. Annotations of the inlets are based on statistically enriched pathways for genes within these modules. Terms with Benjamini–Hochberg FDR adjusted *P*-value < 0.05 in the statistical overrepresentation test were considered statistically significant

### *Coordinated sex-specific modulation of an ensemble of snoRNAs*

Altered co-regulation of 38 snoRNAs was observed only in females exposed to alcohol during development. The direction of changes in snoRNA expression was strongly genetic background-dependent, with some DGRP lines showing coordinated up-regulation, some exhibiting coordinated down-regulation, and others showing no change in snoRNA expression (Figure 4.4; Supplementary File  S4.6). The snoRNAs that exhibit genetic variation in their response to developmental ethanol exposure belong primarily to the H/ACA class, which are associated with pseudouridylation of ribosomal RNAs [29, 30]; many are in introns of *RpS4*, *RpL5*, *RpS5a*, *RpS7*, *RpL11*, *RpS16*, *RpL17*, and *RpL22* ribosomal protein encoding genes. The number of *snoRNA*s within each gene varies from 1 to 15 (Supplementary File  S4.6). When multiple *snoRNA*s are present in a gene, only some show altered expression in response to ethanol exposure, although in some cases clusters of genes, likely expressed as polycistronic transcripts, are regulated together. Other snoRNAs with altered expression in response to chronic ethanol exposure are in introns in *dom*, which contributes to histone acetyl transferase activity associated with epigenetic modification of gene expression [31]; *CG13900* (*Sf3b3*), inferred to form part of a spliceosome complex [32]; *SC35*, which encodes a splicing factor [33]; *kra*, which is annotated as a translation initiation factor binding protein [34]; *Aladin*, predicted to form part of the nucleopore complex [35]; and *Nop60B*, which encodes pseudouridine synthase [36, 37].

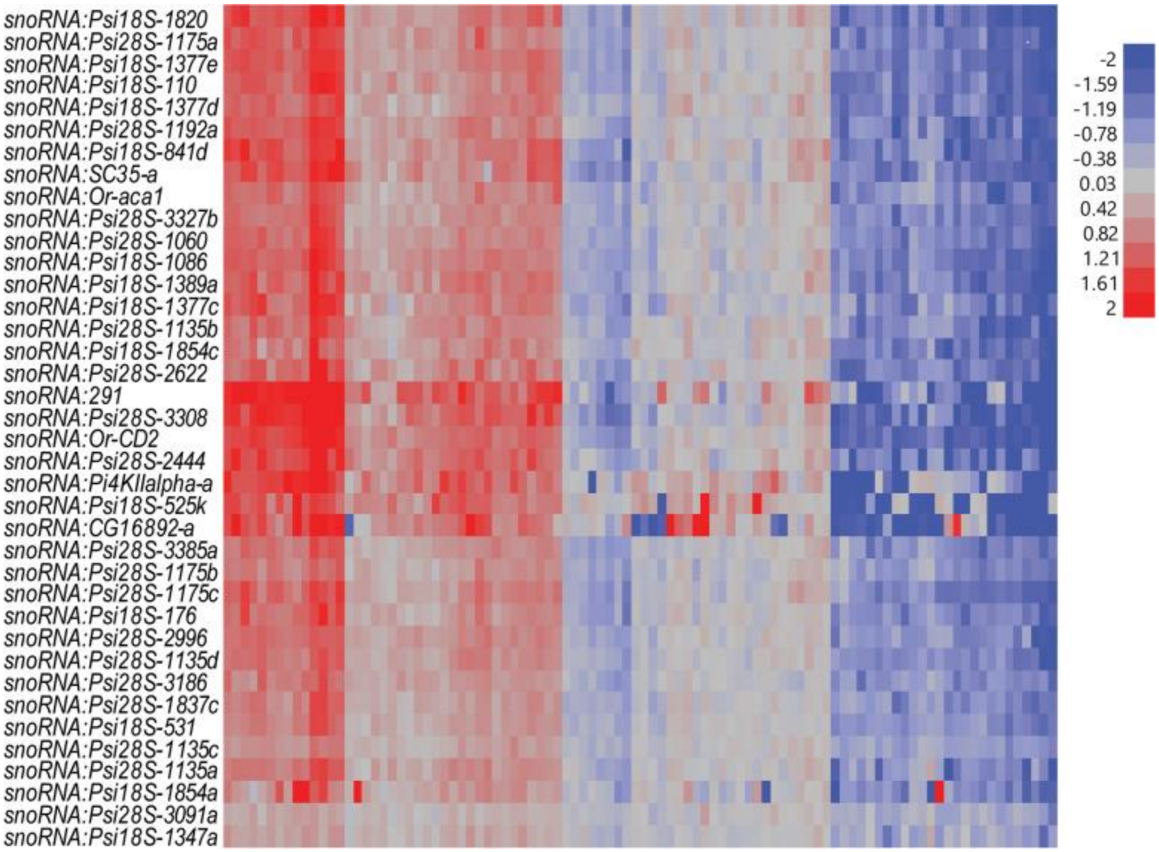


**Figure 4.4. Differentially expressed snoRNAs from female flies grown on ethanol versus regular food.** Vertical columns represent individual DGRP lines. The color scale indicates upregulation (red) or down-regulation (blue) after growth on ethanol-supplemented medium

Examination of variation in transcript abundances of snoRNAs and host genes other than those encoding ribosomal proteins show that the expression of these snoRNAs is regulated independently from the host genes (Figure S4.1). Interestingly, transcript abundance levels of all alcohol-sensitive snoRNAs were correlated with variation in expression of *Uhg4*, which is highly expressed in ovaries and encodes a lncRNA (Figure

S4.1) [38]. A previous genome-wide association study in the DGRP identified *Cyclin E* (*CycE*) as a highly interconnected hub gene in a genetic network associated with alcohol-induced variation in viability and development time [16]. Variation in *CycE* transcript abundance was not correlated with variation in transcript abundance levels of snoRNAs but was highly correlated with variation in transcript abundance levels of their host genes (Figure S4.2). Expression levels of *stet*, *dom*, *Aladin*, *kra* and *Nop60B* were also highly correlated (Figure S4.2).

### *Effects of developmental alcohol exposure on activity and sleep*

The transcriptome is a proximal determinant of organismal phenotypes. FAS/FASD symptoms include hyperactivity and sleep disorders. Since activity and sleep are universal measures of nervous system function which can be modeled in Drosophila, we used activity and sleep parameters as a read-out of the behavioral effects of alcohol exposure during development (Figure 4.5). We selected three DGRP lines which exhibited the highest degree of coordinated up-regulation (DGRP_177, DGRP_208, DGRP_367) and down-regulation (DGRP_555, DGRP_705, DGRP_730) of a subset of snoRNAs in response to developmental alcohol exposure and reared them either on regular food or food supplemented with ethanol. We measured their activity and sleep phenotypes using the Drosophila Activity Monitor (DAM) system, in which single flies are introduced into narrow tubes and a movement is recorded any time the fly disrupts an infrared beam. We performed factorial mixed model ANOVA analyses with the main effects of sex, treatment and DGRP line for activity, night and day sleep proportion, and night and day bout count. There was significant genetic variation for all traits among the DGRP lines, and the *L* × *T* interaction was significant for activity, night bout count, and the proportion of day and night sleep (Supplementary File S4.7). Thus, exposing Drosophila to alcohol during development affects activity and sleep phenotypes relevant to patients with FASD, but the effects are dependent on genetic context.

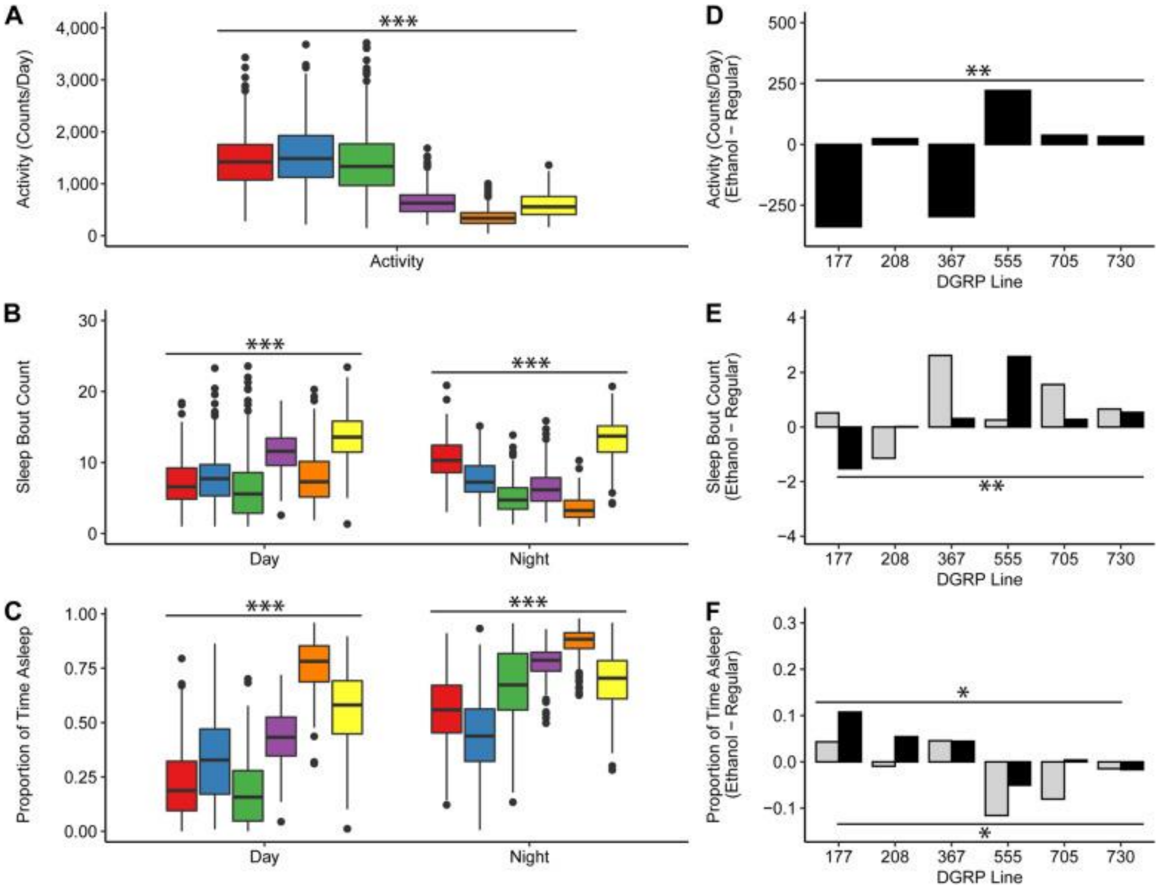


**Figure 4.5 Effects of developmental ethanol exposure on sleep and activity phenotypes.** Boxplots averaged across treatment and sex showing the main effect of line on (**A**) locomotor activity, recorded as the average number of counts per day, where counts are the number of times the fly crosses the infrared beam as recorded by the DAM

System, (**B**) number of sleep bouts during day and night hours, and (**C**) proportion of time spent asleep during day and night hours. DGRP lines 177, 208, 367, 555, 705, 730 are shown in red, blue, green, purple, orange, and yellow, respectively. Bar graphs of (**D**) locomotor activity ($LxT$ $P = 0.0091$), (**E**) number of sleep bouts during day (light grey bars) and night (black bars) hours (Day $LxT$ $P = 0.1538$; Night $LxT$ $P = 0.0014$), and (**F**) proportion of time asleep during day (light grey bars) and night (black bars) hours (Day $LxT$ $P = 0.0491$; Night $LxT$ $P = 0.0181$), for each DGRP line, averaged across treatment, showing the effect of line by treatment (ethanol-supplemented food minus regular food). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

## Discussion

Although previous studies have identified a plethora of genetic risk factors that contribute to alcohol related phenotypes in human populations or rodent models [11, 39], our understanding of the interaction between environmental alcohol exposure and allelic variants remains incomplete. Evolutionary conservation of fundamental biological processes and similarity of the effects of alcohol exposure between flies and people have established *D. melanogaster* as a useful translational gene discovery system [11, 23, 40]. Here, we identified transcripts that undergo altered regulation when flies are reared on ethanol. Changes in transcript abundance patterns are sexually dimorphic and reveal regulatory networks in which NTRs feature prominently. It is of special interest that females exposed to developmental ethanol exposure show altered regulation of a large ensemble of H/ACA class snoRNA genes, which may mediate widespread changes in protein synthesis upon chronic alcohol exposure. The complex relationship between expression of host genes and embedded snoRNAs, which we observe in our studies, has also been documented across seven human tissues, including brain [41]. The mechanisms by which alcohol triggers changes in gene expression remain unknown and could include direct effects on the genome, indirect effects on the genome mediated via alcohol-induced metabolic changes, and/or epigenetic modifications of DNA. Pseudouridylation of mRNAs, tRNAs and other small RNAs in response to environmental stress can consolidate or destabilize interactions between RNAs and proteins [42]. The intimate relationship between snoRNAs and ribosomal function may represent a conduit between the genome and the proteome that can adaptively modulate the composition of the proteome in response to ethanol exposure.

The *D. melanogaster* transcriptome is highly intercorrelated [25] and changes in gene expression due to an environmental disturbance result in modulation of transcriptional niches (*i.e.*, coregulated ensembles) of focal genes [40, 43]. This raises a central "cause *versus* effect" question as it is not evident a priori which gene is the focal gene that directly responds to the environmental perturbation. Insights can be derived from eQTL analyses that can delineate *cis*-eQTLs and *trans*-eQTLs [25, 26]. Interestingly, all but one of the eQTLs associated with transcripts with genetic variation in the difference in expression between standard rearing conditions and developmental exposure to ethanol were in *trans* to the focal transcripts. However, we were able to incorporate the genes to which these eQTLs are *cis* into known interaction networks to derive sex-specific networks associated with genetic variation in response to developmental exposure to ethanol in which these genes are candidate regulatory drivers.

The data we present were obtained from flies that were continuously exposed to alcohol from egg to adult. Exposures that are restricted to different developmental stages will provide a finer grained picture of the dynamics of the alcohol-sensitive genome. Similarly, we analyzed transcriptional responses in whole flies. Single cell RNA sequencing experiments with defined tissues, such as the brain, can provide tissue-specific resolution of the transcriptional response to developmental alcohol exposure [44]. However, the data we obtained in this study underscore extensive sexual dimorphism and emphasize the importance of non-coding elements in regulating the transcriptional response to alcohol exposure during development. Not all aspects of FASD (e.g. cognitive impairment) can be readily modeled in flies. Nevertheless, results from this study illustrate the power of the Drosophila model as a gene discovery system to gain insights into human disorders, such as FASD, that can only be addressed through comparative genomics approaches.

Because of conservation of fundamental biological processes and homologies between fly gene products and their human counterparts, studies on the transcriptional response to developmental exposure to alcohol in *Drosophila melanogaster* can provide insights in the genetic underpinnings that may predispose to FASD.

## Methods

### *Drosophila lines*

The DGRP lines have been generated and are maintained in our laboratories. We selected 96 DGRP lines [20, 21] across the range of phenotypic variation of effects of alcohol exposure on viability and developmental time (Supplementary File S4.1) [16, 24] and reared them on cornmeal-molasses-agar medium supplemented with10% (v/v) ethanol at 25 °C, 60–75% relative humidity and a 12-h light–dark cycle at equal larval densities. We collected two replicates of mated 3–5-day old flies (25 females and 30 males per line) for a total of 384 samples, following procedures described previously for baseline sample collection [26]. We used a randomized experimental design for sample collection that was done strictly between 1–3 pm and froze collected flies over ice supplemented with liquid nitrogen. The flies were sexed and stored in 2.0 ml nuclease-free microcentrifuge tubes (Ambion) at -80 °C until processing.

### *RNA sequencing*

We extracted total RNA as described previously [26] with Trizol using the RNeasy Mini Kit (Qiagen, Inc.), and depleted ribosomal RNA from 5 μg of total RNA using the Ribo-Zero™ Gold Kit (Illumina, Inc.). Depleted mRNA was fragmented and converted to first strand cDNA. During the synthesis of second strand cDNA, we used dUTP instead of dTTP to label the second strand cDNA. We used cDNA from each RNA sample to produce barcoded cDNA libraries using NEXTflex™ DNA Barcodes (Bioo Scientific, Inc.) with an Illumina TruSeq compatible protocol. Libraries were size selected for 250 bp (insert size ~ 130 bp) using Agencourt Ampure XP Beads (Beckman Coulter, Inc.). Second strand DNA was digested with Uracil-DNA Glycosylase before amplification to produce directional cDNA libraries. We quantified the libraries using Qubit dsDNA HS Kits (Life Technologies, Inc.) and Bioanalyzer (Agilent Technologies, Inc.) to calculate molarity. The libraries were subsequently diluted to equal molarity and re-quantified. Samples were processed in batches of 48 and 16 libraries were pooled randomly into 25 pools. We quantified the pooled library samples again to calculate final molarity and after denaturation diluted them to 14 pM. Pooled library samples were clustered on an Illumina cBot; each pool was sequenced on one lane of an Illumina Hiseq2500 using 125 bp single-read v4 chemistry.

### *RNA sequence analysis*

Sequences were analyzed exactly as described previously [26]. We demultiplexed barcoded sequence reads using the Illumina pipeline v1.9 and trimmed adapter sequences using cutadapt v1.6 [45]. The trimmed sequences were aligned to multiple target sequence databases, using BWA v0.7.10 (MEM algorithm with parameters '-v 2 –t 4') [46]. First, we aligned all trimmed sequences against a ribosomal RNA database to filter out residual rRNA that escaped depletion during library preparation. Next, we aligned the remaining sequences against a custom database of potential microbiome component species using BWA. We then aligned sequences that did not align to either the rRNA or microbiome databases to all *D. melanogaster* sequences in RepBase [47]. Finally, we aligned the remaining sequences that did not align to any of the databases above to the *D. melanogaster* genome (BDGP5) and known transcriptome (FlyBase v5.57) using STAR v2.4.0e [48].

*Gene expression estimation*

We followed the analysis described previously [26] to compute read counts for known and novel gene models using HTSeq-count [49] with the 'intersection-nonempty' assignment method. Tabulated read counts for each endogenous gene present in both Baseline and ethanol-treated lines were combined and normalized across all samples using EdgeR [50]. We used the normalized gene expression in all following analyses.

*Genetics of gene expression*

For each expression feature (known and novel transcripts) we fit mixed-effect models to the normalized gene expression data corresponding to: $Y = S + W + T + L + W \times S + L \times S + L \times T + T \times S + T \times S \times L + \varepsilon$, where $Y$ is the observed log2 (normalized read count), $S$ is sex, $W$ is Wolbachia infection status, $W \times S$ is Wolbachia by sex interaction, $L$ is DGRP line, $T$ is treatment (ethanol-supplemented *vs* standard medium), $L \times S$ is the line by sex interaction, $L \times T$ is the line by treatment interaction, $T \times S \times L$ is the treatment by line by sex interaction and $\varepsilon$ is the residual error. We also performed reduced analyses for sexes separately ($Y = W + L + T + L \times T + \varepsilon$).

We identified genetically variable transcripts as those that passed a 5% FDR threshold (based on Benjamini–Hochberg corrected $P$-values [51]) for the $L$, $T$ and $L \times T$ terms. We computed the broad sense heritabilities ($H^2$) for each gene expression trait separately for males and females as

$$H2=\sigma2L+\sigma2SxL+\sigma2TxL+\sigma2SxTxL/(\sigma2L+\sigma2SxL+\sigma2TxL+\sigma2SxTxL+\sigma2\varepsilon),$$

where $\sigma2L$, $\sigma2SxL$, $\sigma2TxL$, $\sigma2SxTxL$ and $\sigma2\varepsilon$ are, respectively, the among line, sex by line, treatment by line, sex by treatment by line, and within line variance components. In addition, for all expression features that were significant for the $L \times T$ interaction term we re-analyzed data for each Treatment condition separately to identify transcripts with significant changes in expression in one or another condition, or both. We performed reduced analyses for sexes separately ($Y = W + L + \varepsilon$). We also calculated the Line means differences for the matching transcripts that were significant for the $L \times T$ term (*i.e.*, Line.ETOH – Line.Baseline) for males and females separately. These line means differences in expression values were used for hierarchical clustering analysis for the subset of the significant gene expression features using the JMP12 package (SAS, Cary, USA).

*Construction of correlated expression networks*

We used the differences between the conditional means with and without ethanol for each line to calculate pairwise Pearson correlation coefficients for all genes that had statistically significant Benjamini–Hochberg False Discovery Rate (BH-FDR) [51] adjusted p-values (BH-FDR $< 0.05$) for the line-by-ethanol treatment interaction term in the linear mixed effects model ($Y = W + L + T + LxT + \varepsilon$ run for each sex separately). The delta-expression correlation matrix was bi-clustered using hierarchical clustering with complete linkage agglomeration in Genesis statistical software [52]. The bi-clustered matrix was used for the center panel in each composite figure (Figures 4.2 and 4.3). To generate the delta-expression correlation networks, the correlation matrix was filtered for associations with BH-FDR adjusted $p$-value $< 0.05$ and the top 10% of correlations based on the absolute value of the Pearson coefficient. Associations that survive the stringent filtering criteria were input into Cytoscape and clustered using the MCODE algorithm with default parameters, but with the 'Fluff' setting activated to capture relationships outside of auto-correlated modules [53]. The resulting modules with significantly strong intra-connectivity (cumulative MCODE score $> 4$) were mapped back to the correlation matrix panel based on the identity of the gene membership of each module. We highlighted genes with a statistically significant (BH-FDR $< 0.05$) eQTL association calculated from the expression differences in each module within the composite figure.

### *eQTL mapping*

eQTLs were mapped to differences in expression between baseline and ethanol treatment of genes with a statistically significant line-by-treatment ($LxT$) term from the linear mixed effects model run for each sex separately as previous described [25, 26]. Briefly, we adjusted normalized FPKM values for *Wolbachia* infection status, chromosomal inversions, population structures organized based on top 10 principal components using Best Linear Unbiased Predictor (BLUP) using the R package *lmerTest*. We used covariate adjusted expression differences as phenotype for eQTL mapping using PLINK (v1.90). We compared association *P*-values generated by the PLINK *t*-tests to the empirical FDR threshold calculated by dividing the number of expected associations under the null hypothesis generated from 100 permutations at a false discovery rate of 0.05 by the observed number of associations at the same threshold to determine statistical significance. We further filtered associations filtered for independence using forward model selection, as previously described [25, 26], by iteratively adding single eQTLs, starting with the smallest *P*-values, to an additive association model such that the conditional *P*-value for the last added eQTL is no more than 1E-5.

### *Association networks*

To build association networks using the variants identified from eQTL mapping, we added pairwise eQTL associations between genes that either contain the variant or are within 1000 bp up- or down-stream of the variant and genes with statistically significant line-by-treatment (*LxT*) to the most recent version (fb_2021_05) of the database of known genetic and protein–protein or RNA–protein physical interactions from the FlyBase repository and visualized in Cytoscape. The resulting networks of associations were filtered to contain (i) genes that are part of the eQTL associations, and (ii) genes that have at least 5 genetic or physical interactions to genes that are part of the eQTL associations within one interaction distance (one edge). The filtered interaction network was modularized using MCODE algorithm with default settings but with ‘fluff’ activated [53]. We input genes that are part of the individual modules for Gene Ontology

Enrichment analysis. Statistically significant (FDR < 0.05 in the Overrepresentation Test), highly specialized terms containing the largest number of genes from the input from each module were used for functionally labeling each module. For the glutathione metabolism inlet in the female interaction network in Figure 4.3, *GstE* family of genes were added to the *GstD* tri-gene cluster based on semantic similarity.

### *Analysis of activity and sleep phenotypes*

Flies were reared on standard medium and medium supplemented with 10% (v/v) ethanol and placed in standard food collection vials overnight. The next day, mated females and males from both treatment conditions were placed in Drosophila Activity Monitor (DAM) tubes (TriKinetics, Waltham, MA) that contained agar supplemented with 5% sucrose at one end, and a small piece of yarn at the other end. Flies were placed in a 25 °C incubator on a 12-h light–dark cycle and their activity and sleep data were recorded using the DAMSystem (Trikinetics). Raw data from the DAMSystem (TriKinetics) were uploaded to ShinyR-DAM [54] and resulting output data were parsed by sleep/activity phenotype for analysis. Sleep was defined as at least 5 min of inactivity and only data from flies that survived the entire testing period (2–9 days of the fly lifespan) were retained for analysis. Sleep and activity data were analyzed using the PROC MIXED command (Type III Analysis of Variance (ANOVA)) within SAS version 9.04 (Cary, NC) according to the model $Y = \mu + T + L + S + TxL + TxS + LxS + TxLxS + Rep(TxLxS) + \varepsilon$, where $T$ is the fixed effect of treatment (ethanol medium, standard medium), $L$ is the fixed effect of line (RAL_177, RAL_208, RAL_367, RAL_555, RAL_705, RAL_730), *S* is the fixed effect of Sex (male, female), *Rep(TxLxS)* is the random effect of replicate, and $\varepsilon$ is the residual variance. Reduced Type III ANOVAs ($Y = \mu + T + L + TxL + Rep(TxL) + \varepsilon$) were also performed by Sex.

DGRP lines are available from the Bloomington Drosophila Stock Center, https://bdsc.indiana.edu/. RNA sequencing data have been deposited in the GEO database https://www.ncbi.nlm.nih.gov/geo/ under accession number GSE186240. Raw DAM data are on the github repository: https://github.com/rebeccamacpherson/DAM_raw_data_DEV_ETOH_DGRP

## Acknowledgements

We thank Gunjan Arya, Lavanya Turlapati and Genevieve St. Armour for technical assistance.

This work was supported by grants from the National Institutes on Alcohol and Alcoholism (AA016560) and the National Institute on Drug Abuse (DA041613) to RRHA and TFCM and by a Ruth L. Kirschstein National Research Service Award from the National Institute of Child Health and Human Development (HD106719) to RAM. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

# References

1. Jones K, Smith D. Recognition of the fetal alcohol syndrome in early infancy. The Lancet. 1973;302: 999-1001.

2. Hoyme HE, May PA, Kalberg WO, Kodituwakku P, Gossage JP, Trujillo PM, et al. A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: clarification of the 1996 Institute of Medicine criteria. Pediatrics. 2005;115:39-47.

3. Roozen S, Peters GJ, Kok G, Townend D, Nijhuis J, Curfs L. Worldwide prevalence of Fetal Alcohol Spectrum Disorders: A systematic literature review including meta-analysis. Alcohol Clin Exp Res. 2016:40:18-32.

4. Kaminen-Ahola N. Fetal alcohol spectrum disorders: Genetic and epigenetic mechanisms. Prenat Diagn. 2020;40:1185-92.

5. Clarren SK, Smith DW. The fetal alcohol syndrome. Lamp. 1978;35:4-7.

6. Pulsifer MB. The neuropsychology of mental retardation. J Int Neuropsychol Soc. 1996;2:159-76.

7. Eckardt MJ, File SE, Gessa GL, Grant KA, Guerri C, Hoffman PL, et al. Effects of moderate alcohol consumption on the central nervous system. Alcohol Clin Exp Res. 1998:22:998-1040.

8. Lange S, Rehm J, Anagnostou E, Popova S. Prevalence of externalizing disorders and Autism Spectrum Disorders among children with Fetal Alcohol Spectrum Disorder: systematic review and meta-analysis. Biochem Cell Biol. 2018;96;241-51.

9. Tan CH, Denny CH, Cheal NE, Sniezek JE, Kanny D. Alcohol use and binge drinking among women of childbearing age - United States, 2011-2013. MMWR. 2015;64:1042- 46.

10. Petrelli B, Weinberg J, Hicks GG. Effects of prenatal alcohol exposure (PAE): insights into FASD using mouse models of PAE. *Biochem Cell Biol.* 2018:96:131-47.

11. Morozova TV, Goldman D, Mackay TFC, Anholt RRH. The genetic basis of alcoholism: multiple phenotypes, many genes, complex networks. Genome Biol. 2012;13:239.

12. Morozova TV, Mackay TFC, Anholt RRH. Genetics and genomics of alcohol sensitivity. Mol Genet Genomics. 2014;289:253-69.

13. Scholz H. Unraveling the mechanisms of behaviors associated with AUDs using flies and worms. Alcohol Clin Exp Res. 2019;43:2274-84.

14. Petruccelli E, Kaun KR. Insights from intoxicated Drosophila. Alcohol. 2019;74:21-7.

15. McClure KD, French RL, Heberlein UA. Drosophila model for fetal alcohol syndrome disorders: role for the insulin pathway. Dis Model Mech. 2011;4:335-46.

16. Morozova TV, Hussain Y, McCoy LJ, Zhirnov EV, Davis MR, Pray VA, et al. A *Cyclin E* centered genetic network contributes to alcohol-induced variation in Drosophila development. G3 (Bethesda). 2018;8:2643-53.

17. Scepanovic G, Stewart BA. Analysis of Drosophila nervous system development following an early, brief exposure to ethanol. Dev Neurobiol. 2019;79:780-93.

18. Logan-Garbisch T, Bortolazzo A, Luu P, Ford A, Do D, Khodabakhshi P, French RL. Developmental ethanol exposure leads to dysregulation of lipid metabolism and oxidative stress in Drosophila. G3 (Bethesda). 2014:5:49-59.

19. Guevara A, Gates H, Urbina B, French R. Developmental ethanol exposure causes reduced feeding and reveals a critical role for Neuropeptide F in survival. Front Physiol. 2018;9:237.

20. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The *Drosophila melanogaster* Genetic Reference Panel. Nature. 2012;482:173-8.

21. Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, et al. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. Genome Res. 2014;24:1193-1208.

22. Fochler S, Morozova TV, Davis MR, Gearhart AW, Huang W, Mackay TFC, et al. Genetics of alcohol consumption in *Drosophila melanogaster*. Genes Brain Behav. 2017;16:675-85.

23. Morozova TV, Ayroles JF, Jordan KW, Duncan LH, Carbone MA, Lyman RF, et al. Alcohol sensitivity in Drosophila: translational potential of systems genetics. Genetics. 2009;183:733-45.

24. Morozova TV, Huang W, Pray VA, Whitham T, Anholt RRH, Mackay TFC. Polymorphisms in early neurodevelopmental genes affect natural variation in alcohol sensitivity in adult drosophila. BMC Genomics. 2015;16:865.

25. Huang W, Carbone MA, Magwire MM, Peiffer JA, Lyman RF, Stone EA, et al. Genetic basis of transcriptome diversity in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 2015;112:E6010-9.

26. Everett LJ, Huang W, Zhou S, Carbone MA, Lyman RF, Arya GH, et al. Gene expression networks in the Drosophila Genetic Reference Panel. Genome Res. 2020;30:485-96.

27. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, et al. Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet. 2009;41:299-307.

28. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc. 2013;8:1551-66.

29. Reichow SL, Hamma T, Ferré-D'Amaré AR, Varani G. The structure and function of small nucleolar ribonucleoproteins. Nucleic Acids Res. 2007;35:1452-64.

30. Yu YT, Meier UT. RNA-guided isomerization of uridine to pseudouridine--pseudouridylation. RNA Biol. 2014:11:1483-94.

31. Kusch T, Florens L, Macdonald WH, Swanson SK, Glaser RL, Yates JR 3rd, et al. Acetylation by *Tip60* Is required for selective histone variant exchange at DNA lesions. Science. 2004;306:2084-7.

32. Stegeman R, Spreacker PJ, Swanson SK, Stephenson R, Florens L, Washburn MP, et al. The spliceosomal protein SF3B5 is a novel component of Drosophila SAGA that functions in gene expression independent of splicing. J Mol Biol. 2016;428:3632-49.

33. Bradley T, Cook ME, Blanchett M. SR proteins control a complex network of RNA-processing events. RNA. 2015;21:75-92.

34. Lee S, Nahm M, Lee M, Kwon M, Kim E, Zadeh AD, et al. The F-actin-microtubule crosslinker Shot is a platform for Krasavietz-mediated translational regulation of midline axon repulsion. Development. 2007;134:1767-77.

35. Carvalhal S, Ribeiro SA, Arocena M, Kasciukovic T, Temme A, Koehler K, et al. The nucleoporin ALADIN regulates Aurora A localization to ensure robust mitotic spindle formation. Mol Biol Cell. 2015;26:3424-38.

36. Giordano E, Peluso I, Senger S, Furia M. *minifly*, a Drosophila gene required for ribosome biogenesis. J Cell Biol. 1999;144:1123-33.

37. Riccardo S, Tortoriello G, Giordano E, Turano M, Furia M. The coding/non-coding overlapping architecture of the gene encoding the Drosophila pseudouridine synthase. BMC Mol Biol. 2007;8:15.

38. Herter EK, Stauch M, Gallant M, Wolf E, Raabe T, Gallant P. snoRNAs are a novel class of biologically relevant *Myc* targets. BMC Biol. 2015;13:25.

39. Warden AS, Mayfield RD. Gene expression profiling in the human alcoholic brain. Neuropharmacol. 2017;122:161-74.

40. Anholt RRH, Mackay TFC. The road less traveled: From genotype to phenotype in flies and humans. Mamm Genome. 2018;29:5-23.

41. Fafard-Couture E, Bergeron D, Couture S, Abou-Elela S, Scott MS. Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships. Genome Biol. 2021:22:172.

42. Borchardt EK, Martinez NM, Gilbert WV. Regulation and function of RNA pseudouridylation in human cells. Annu Rev Genet. 2020:54:309-36.

43. Anholt RRH, Dilda CL, Chang S, Fanara JJ, Kulkarni NH, Ganguly I, et al. The genetic architecture of odor-guided behavior in Drosophila: Epistasis and the transcriptome. Nat Genet. 2003;35:180-4.

44. Mokashi SS, Shankar V, MacPherson RA, Hannah RC, Mackay TFC, Anholt RRH. Developmental alcohol exposure in Drosophila: Effects on adult phenotypes and gene expression in the brain. Front Psychiatry. 2021;12:699033.

45. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10-2.

46. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589-95.

47. Jurka J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462-7.

48. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15-21.

49. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166-9.

50. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics**.** 2010;26;139-40.

51. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B (Methodological). 1995;57:289-300.

52. Sturn A, Quackenbush J, Trajanoski Z. Genesis: Cluster analysis of microarray data. Bioinformatics. 2002:18:207-8.

53. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2.

54. Cichewicz K, Hirsh J. ShinyR-DAM: a program analyzing Drosophila activity, sleep and circadian rhythms. Commun. Biol. 2018;1:25.

# CHAPTER FIVE

# PLEIOTROPIC FITNESS EFFECTS OF THE LNCRNA *UHG4* IN *DROSOPHILA MELANOGASTER*

MacPherson RA, Shankar V, Sunkara LT, Hannah RC, Campbell MR 3rd, Anholt RRH, Mackay TFC. Pleiotropic fitness effects of the lncRNA *Uhg4* in *Drosophila melanogaster*. *BMC Genomics*. 2022 Nov 30;23(1):781. doi: 10.1186/s12864-022-08972-0.

## Author Contribution Statement

My contributions to the work described in this chapter are as follows: With the assistance of Rachel Hannah and Marion Campbell, I collected flies and performed behavioral experiments. I designed all of the experiments, created the gene editing constructs, and performed the Sanger sequencing, dissections, and staining. I also performed all analysis for the behavioral data. I performed the sequencing analysis and interpretation with the assistance of Vijay Shankar. I also wrote the original draft of the manuscript.

## Introduction

Long noncoding RNAs (lncRNAs) are a diverse class of non-coding RNAs of at least 200 nucleotides in length. Although lncRNAs were initially thought to be “junk” due to their noncoding status, we now know that lncRNAs are critical for various biological processes, including, but not limited to, transcription and gene regulation [1–8], chromatin architecture [9–11], DNA damage response [12–14], and scaffolding and nuclear organization [15–18]. LncRNAs can also act as miRNA sponges [19–21], regulate gene splicing [17, 22–25], and be translated into functional peptides [26–28]. LncRNAs can be localized in the nucleus, cytoplasm, or to a specific organelle [29–32], can act in *cis* or in *trans* [8, 23, 33], and may be conserved across taxa in sequence and/or function [34–36]. Across taxa, lncRNA dysregulation has been implicated in cancer [37–40], neurological disorders [6, 41–43], and immune and stress response [44–47]. Additional studies in *Drosophila melanogaster* have demonstrated critical roles for lncRNAs in development [48–53], gonadal function [48, 54–56], sleep [19], locomotion [57], and courtship behavior [58, 59].

Roles for lncRNAs in development, viability, and fertility have been identified in multiple model systems [4, 8, 34, 55, 60, 61]. However, fitness roles for some mammalian lncRNAs, including lncRNAs previously linked to fitness traits with gene-knockdown experiments, have not been replicated using CRISPR-based cell line or animal model knockouts [62–65]. Other recently discovered mammalian lncRNAs with expression limited to reproductive tissues do not affect reproductive phenotypes in knockout mice [66, 67]. This controversy also extends beyond mammalian systems. Drosophila and zebrafish studies on CRISPR-mediated deletions of developmentally-expressed lncRNAs – some of which were previously implicated in development via RNA interference or morpholino-induced knockdown studies – also failed to identify roles for these lncRNAs in development, viability, or embryogenesis after no overt phenotypes were detected in knockout animals [68–70]. Inconsistencies across lncRNA studies may be due to the method of gene perturbation (*e.g.*, RNAi interference, CRISPR-*Cas9*), discordant phenotyping, transcriptional noise competing with low expression of some lncRNAs, differences in the genetic background of organisms used across studies on the same lncRNA, and/or functional redundancy of the lncRNA [61, 69, 71, 72]. Furthermore, given the abundance of sex-specific and tissue-specific lncRNA expression, it is also possible that broader developmental pathways may overshadow effects due to loss of a lncRNA or be limited to a single tissue or behavioral phenotype [61, 72].

Here, we evaluate the effects of loss of function alleles of the *D. melanogaster* gene encoding *U snoRNA host gene 4* (*Uhg4*; FBgn0083124). *Uhg4* is a lncRNA with unknown function and is the host gene for seven intronic small nucleolar RNAs (snoRNAs), which guide posttranscriptional ribosomal RNA modification and processing [73, 74]. Some *U snoRNA* host genes interact with regulatory proteins controlling *piwi*-interacting RNAs (piRNAs) [75], which are noncoding RNAs involved in transposon silencing in germ cells [76, 77]. Although adult expression of *Uhg4* is highest in ovaries, it is expressed ubiquitously during development and in other adult tissues, including the brain and central nervous system, fat body, and trachea [78, 79]. *Uhg4* expression is correlated with modulation of expression of a subset of snoRNAs in response to developmental alcohol exposure in Drosophila females [80]. We used CRISPR/*Cas9* germline gene editing to create deletions in the promoter region and first exon of *Uhg4*. These mutations have pleiotropic effects on fitness-related traits, stress responses, sleep and activity phenotypes, and transcript abundances of both non-coding and protein-coding RNAs.

## Results

### *Generation of Uhg4 Null Alleles*

*Uhg4* is a long noncoding RNA that is host to seven small nucleolar RNAs (snoRNAs), and is expressed ubiquitously throughout development and in adults, with the highest adult expression in ovaries [78, 79]. We used CRISPR-*Cas9* and a double guide RNA vector to target the deletion of a ~ 685 bp region that includes the promoter region upstream of *Uhg4* as well as the first exon of *Uhg4* (Figure 5.1) in two lines of the *Drosophila melanogaster* Genetic Reference Panel (DGRP [81]), DGRP lines, DGRP_208 and DGRP_705. We isolated seven independent deletion mutations in DGRP_208, most of which varied from one another by a few base pairs (Figure 5.1). The DGRP_208 deletions spanned the promoter region and the first exon of *Uhg4*. We also isolated four independent identical mutations in DGRP_705. The DGRP_705 deletions removed the *Uhg4* promotor region and retained the first exon, and included a 44 bp AT-rich insertion in the first *Uhg4* intron upstream of the start of *snoRNA:Psi28S-2949* (Figure 5.1). Sanger sequencing shows that the small nucleolar RNA (snoRNA) *snoRNA:Psi28S-2949* within the first exon of *Uhg4* is intact for all independently obtained deletions (Figure 5.1).

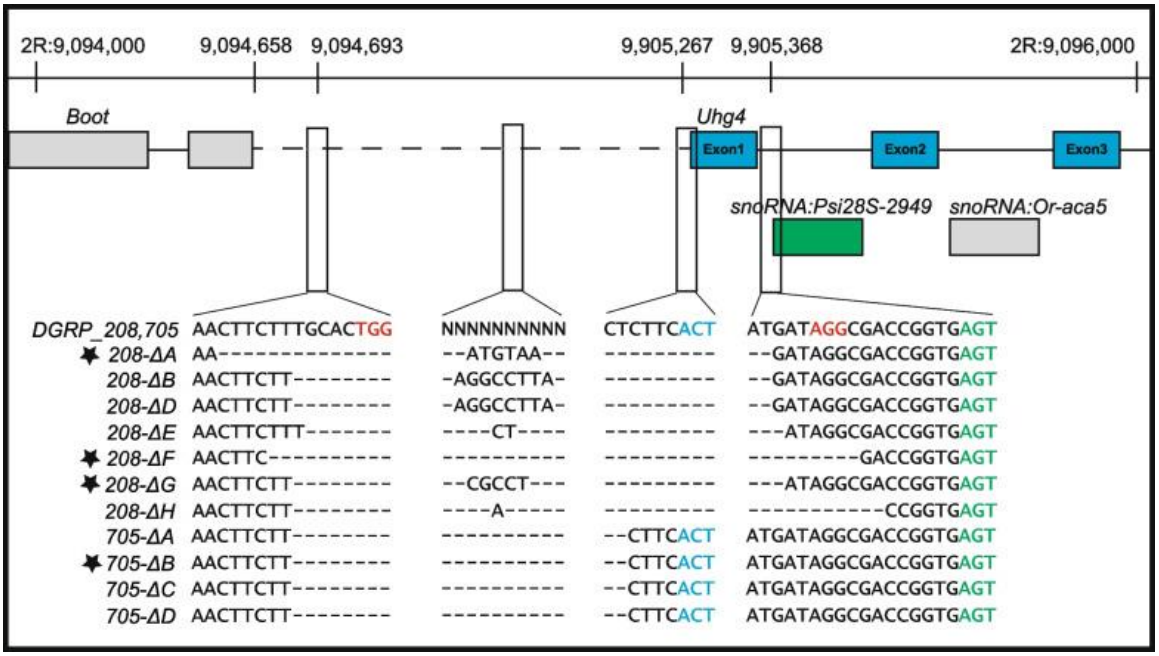

2R:9,094,000
9,094,658
9,094,693
9,905,267
9,905,368
2R:9,096,000
Boot
Uhg4
Exon1
Exon2
Exon3
snoRNA:Psi28S-2949
snoRNA:Or-aca5
DGRP_208,705 AACTTCTTTGCACTGG NNNNNNNNNN CTCTTCACT ATGATAGGCGACCGGTGAGT
208-ΔA AA-------------- --ATGTAA-- --------- --GATAGGCGACCGGTGAGT
208-ΔB AACTTCTT-------- -AGGCCTTA- --------- --GATAGGCGACCGGTGAGT
208-ΔD AACTTCTT-------- -AGGCCTTA- --------- --GATAGGCGACCGGTGAGT
208-ΔE AACTTCTTT------- ----CT---- --------- ---ATAGGCGACCGGTGAGT
208-ΔF AACTTC---------- ---------- --------- ---------GACCGGTGAGT
208-ΔG AACTTCTT-------- --CGCCT--- --------- ---ATAGGCGACCGGTGAGT
208-ΔH AACTTCTT-------- ----A----- --------- -----------CCGGTGAGT
705-ΔA AACTTCTT-------- ---------- --CTTCACT ATGATAGGCGACCGGTGAGT
705-ΔB AACTTCTT-------- ---------- --CTTCACT ATGATAGGCGACCGGTGAGT
705-ΔC AACTTCTT-------- ---------- --CTTCACT ATGATAGGCGACCGGTGAGT
705-ΔD AACTTCTT-------- ---------- --CTTCACT ATGATAGGCGACCGGTGAGT

**Figure 5.1. *Uhg4* deletions.** Diagram showing deletions across the *Uhg4* locus with Sanger sequencing data for the wildtype (DGRP_208, DGRP_705) and mutant genotypes. Genomic coordinates are shown at the top of the figure. The blue and green font colors indicate nucleotides of *Uhg4* and *snoRNA:Psi28S*-2949 coding regions, respectively. Red font colors indicate the PAM sites. "- " indicates deleted nucleotides and N refers to multiple different nucleotides between the two wild-type DGRP lines. Genotypes used for phenotypic analyses are designated with a star

Here, we focus on deletions $Uhg4^{208-\Delta A}$, $Uhg4^{208-\Delta F}$, $Uhg4^{208-\Delta G}$, and $Uhg4^{705-\Delta B}$, hereafter referred to as *208-ΔA*, *208-ΔF*, *208-ΔG,* and *705-ΔB*, respectively. We randomly selected these mutations for further study after several generations of backcrossing to the original genetic background. Both genetic backgrounds are highly inbred and free of inversions, and DGRP_208 is free of *Wolbachia* infection [81]. Flies homozygous for the DGRP_208 *Uhg4* deletions show changes in their resting wing position (Supplementary File S5.1, S5.2). Compared to the control, *Uhg4* deletion flies carry their wings in an elevated and partially horizontally spread position.

### *Effects of Uhg4 Deletions on Fitness Traits*

#### Fertility and Mating Behavior

When we generated the *Uhg4* deletion fly stocks, we did not observe eggs in vials that contained only flies homozygous for a *Uhg4* deletion and we needed to use a *CyO* balancer chromosome to maintain the deletion lines (Supplementary Figure S5.1). To test whether the lack of eggs could be due to a failure of the *Uhg4* deletion flies to mate, we assessed mating latency and copulation duration and found that, although deletion flies do not produce progeny when mated, they do exhibit normal mating latencies, copulation times, and proportion of flies mated (Supplementary Figure S5.2, Supplementary Tables S5.1A, S5.1B). After crossing flies with the deletion to wild-type DGRP_208 flies of the opposite sex, we observed that *Uhg4* deletion females do not lay eggs, regardless of mating status or genotype of the male partner, and wild-type eggs fertilized by *Uhg4* deletion males do not develop past the embryo stage. To further probe why *Uhg4* deletion females are sterile, we dissected their ovaries after mating. Whereas ovaries in control DGRP_208 flies contain late-stage oocytes with dorsal filaments (Fig. 2A), we did not observe any late-stage (> 11) oocytes in ovaries of *Uhg4* deletion females (Fig. 2B and andC).C). We did not perform these analyses for the *705-ΔB*

*Uhg4* allele because it is nearly lethal as a homozygote, with rare viable adults. However, the few escaper *705-ΔB* females did not lay eggs. These results show that *Uhg4* is critical for fertility in both sexes and that the sterility of *Uhg4* deletion in females may stem from a lack of fully developed eggs.

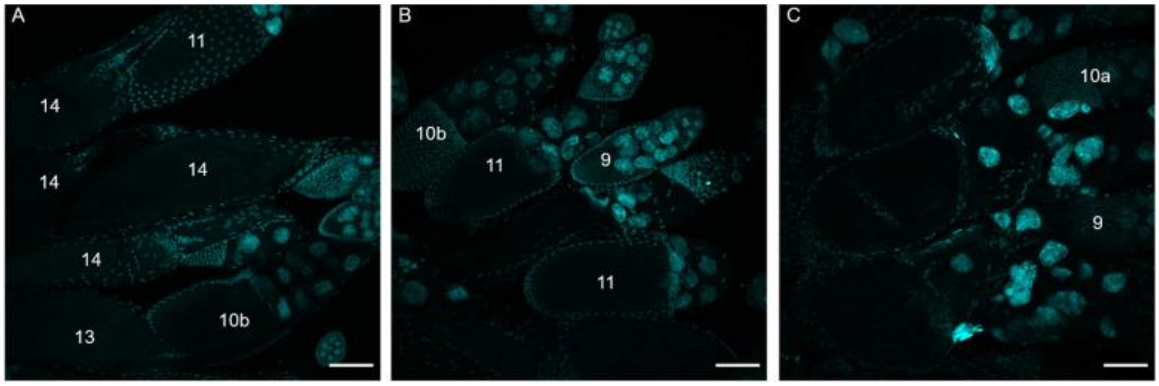


**Figure 5.2. Late-stage oocytes are absent in *Uhg4* deletion lines.** Maximum projection z-stack images (60 slices) of DAPI-stained developing egg chambers. **A** Control *Uhg4* (DGRP_208) ovaries. **B** *208-ΔF* ovaries. **C** *208-ΔG* ovaries. Compared to wild type, late-stage oocytes (stage number > 11) and any dorsal filaments are absent in the ovaries of flies with a *Uhg4* deletion. Scale bars represent 75 µm. Numbers represent stages based on Jia et al. [110]

Development Time and Viability

During maintenance of *CyO*/*Uhg4*-deletion flies, we observed delayed emergence and skewed non-Mendelian ratios of *CyO*/deletion heterozygotes and homozygous deletion progeny. We formally assessed egg-adult development time and viability for the DGRP_208 *Uhg4* deletion lines compared to the wild-type control. We placed 50 eggs from matings of *208-ΔA*/*CyO*, *208-ΔF*/*CyO*, *208-ΔG*/*CyO* or DGRP_208/*CyO* (wild type control) flies in 25 vials each and recorded the day each fly emerged, as well as the sex, balancer genotype (*Cy* or straight wing), and the total number of flies for each sex and balancer genotype that emerged. Compared to the wild-type controls, flies with *Uhg4* deletions have delayed development by about one day ($p < 0.0001$, Figure 5.3A, Supplementary Table S5.1C). Furthermore, *208-ΔA* males and all *Uhg4* deletion females show a 2.0- to 5.7-fold decrease in viability ($p < 0.0001$ for all lines, Figure 5.3B, Supplementary Tables S5.1D, S5.1E). These results suggest that *Uhg4* is necessary for the normal timing of egg-adult development and viability.

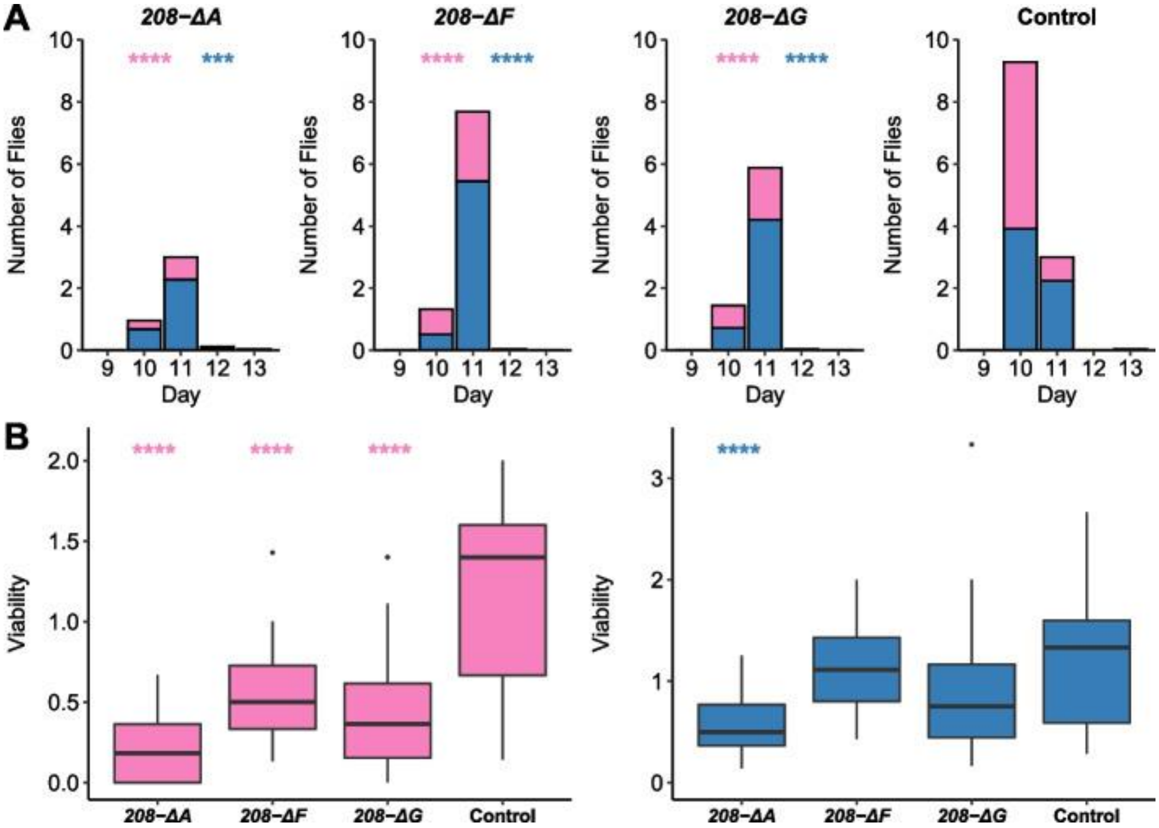


**Figure 5.3. Effects of *Uhg4* deletion on egg-adult development and viability. A** Stacked bar plots showing the average number of flies homozygous for *208-ΔA*, *208-ΔF*, *208-ΔG*, or wild type *Uhg4* (DGRP_208) that emerge from crosses of *208-ΔA*/*CyO*, *208-ΔF*/*CyO*, *208-ΔG*/*CyO* and DGRP_208/*CyO* flies. **B**-**C** Boxplots displaying viability coefficients for females (**B**) and males (**C**) for *Uhg4* deletion lines (*208-ΔA*, *208-ΔF*, *208-ΔG*) and the wild type (DGRP_208). Males are shown in blue and females are shown in pink. $N = 25$ vials of 50 embryos each per genotype. See Table S1 for all ANOVAs. *p*-values on the figure are for the comparisons of each sex to the control, by genotype. *** $p < 0.001$, **** $p < 0.0001$

### *Effects of Uhg4 Deletions on Responses to Stressors and Sleep Traits*

#### Stress Responses

We assessed the effect of several stress conditions (heat shock, chill coma recovery, and ethanol sedation) on *Uhg4* deletion and wild-type control flies (Figure 5.4). On average, *Uhg4* deletion lines take longer to recover from a chill-induced coma than the wild-type in analyses pooled across sexes and all deletion genotypes ($p = 0.013$, Supplementary Table S5.1F), although there is no difference in the proportion of flies that recover from a chill-induced coma for the deletion lines compared to the control (Supplementary Table S5.1G). However, the response to a chill-induced coma is both sex- and genotype-specific. The chill coma recovery time for all deletion lines compared to the wild type is significant for males ($p = 0.0015$) but not females ($p = 0.666$); and males are only significant for the *208-ΔA* ($p < 0.0001$) and *208-ΔF* ($p = 0.0007$) deletion genotypes (Figure 5.4A). The *Uhg4* deletion lines also have reduced survival on average following a heat shock than the wild type in analyses pooled across sexes and all deletion genotypes ($p = 0.0016$, Supplementary Table S5.1F). These effects were genotype-specific as well; only *208-ΔF* ($p = 0.0013$) and *208-ΔG* ($p = 0.034$) were formally significantly different from the wild type, although *208-ΔA* trended in the same direction ($p = 0.093$) (Figure 5.4B). In contrast to temperature-related stressors, *Uhg4* deletion flies show decreased susceptibility to ethanol-induced stress. The time to sedation in response to acute ethanol exposure is significantly increased averaged over all *Uhg4* deletion lines compared to the wild type in the analyses pooled across sexes ($p < 0.0001$) as well as in females ($p = 0.0007$) and males ($p < 0.0001$) (Supplementary Table S5.1F). Although all *Uhg4* deletion sex/genotype comparisons had significantly increased sedation times relative to the control, there was heterogeneity in the magnitudes of effects among the deletion genotypes and sexes. In females, *208-ΔF* ($p < 0.0001$) had larger effects than *208-ΔA* ($p = 0.016$) and *208-ΔG* ($p = 0.036$); while in males all *Uhg4* deletion lines had similar effects ($p < 0.0001$ for all) (Figure 5.4C).

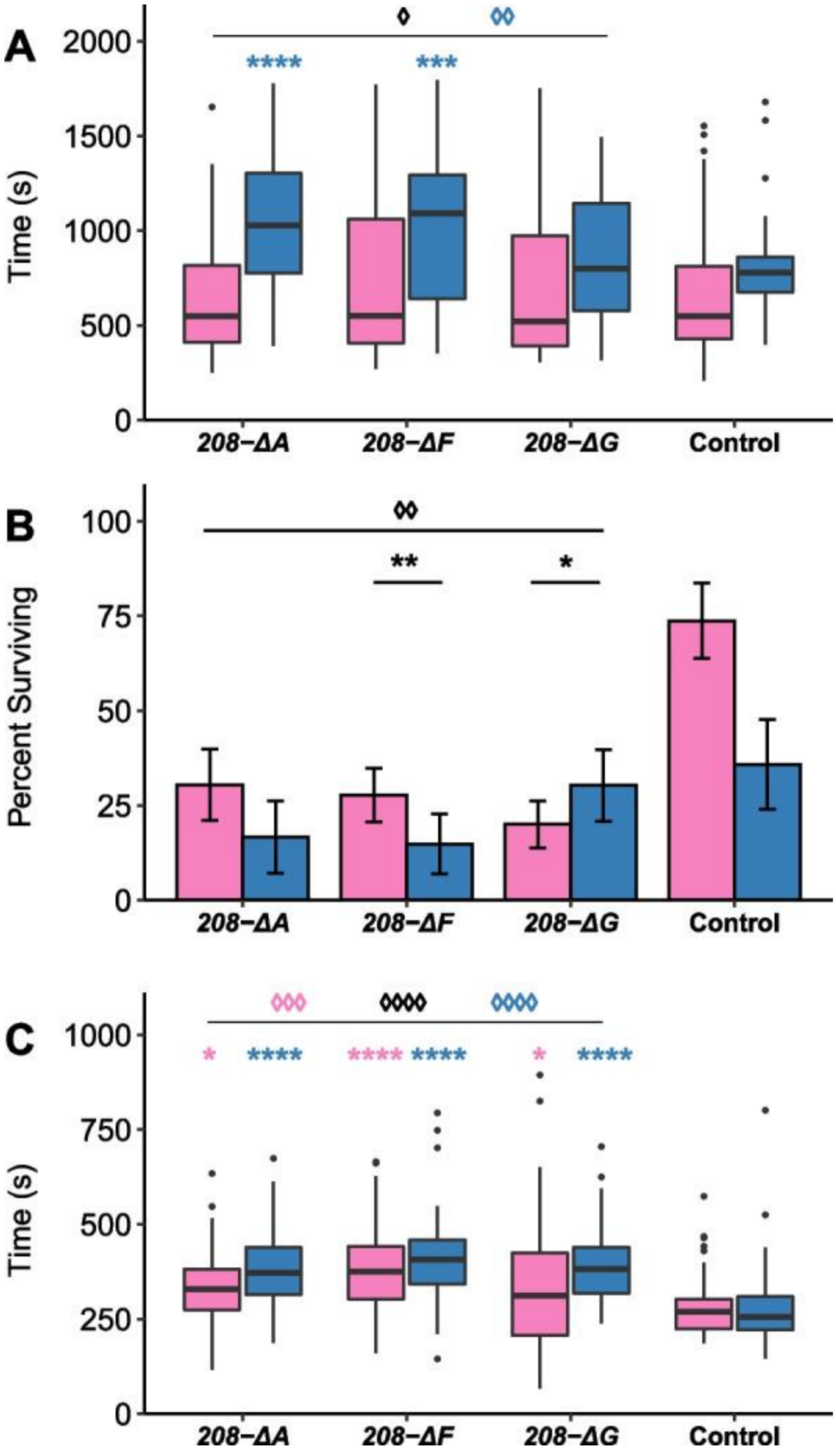

A
Time (s)
2000
1500
1000
500
0
◊
◊◊
****
***
208−ΔA
208−ΔF
208−ΔG
Control
B
Percent Surviving
100
75
50
25
0
◊◊
**
*
208−ΔA
208−ΔF
208−ΔG
Control
C
Time (s)
1000
750
500
250
0
◊◊◊
◊◊◊◊
◊◊◊◊
*
****
****
****
*
****
208−ΔA
208−ΔF
208−ΔG
Control

**Figure 5.4. Effects of *Uhg4* deletion on responses to stress. A** Boxplots indicating chill coma recovery time ($n = 48–57$ flies per sex and genotype). **B** Bar plots showing the average proportion of flies surviving heat shock ($n = 10$ replicates of 9 flies per sex and genotype). **C** Boxplots showing ethanol sedation sensitivity time ($n = 44–52$ flies per sex and genotype). Blue boxes indicate males and pink boxes indicate females. Error bars indicate standard error. See Supplementary Table S5.1 for all ANOVAs. *p*-values on the figure indicated by an asterisk (*) represent comparisons of each sex to the control, by genotype. *p*-values on the figure indicated by a diamond (◊) represent comparisons of pooled *Uhg4* deletion genotypes versus the control. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Sleep and Activity

We used the Drosophila Activity Monitor (DAM) System to assess the effects of *Uhg4* deletions on sleep and activity traits.

*208-ΔA*, *208-ΔF*, and *208-ΔG* flies sleep more during the day ($p = 0.0009$ from the analysis of all deletion lines pooled across sexes compared to the wild type, Figure 5.5A, Supplementary Table S5.1H) and night ($p = 0.031$ from the analysis of all deletion lines pooled across sexes compared to the wild type, Figure 5.5B, Supplementary Table S5.1H). The effects of the deletions are much greater on day than on night sleep. In addition, day sleep is significant averaged over all deletion lines for males ($p = 0.026$) and females ($p = 0.014$); while the effect on night sleep is male-specific ($p = 0.011$)

(Supplementary Table S5.1H). Although the *Uhg4* deletions sleep longer than the wild type during the day and night, they also have more fragmented sleep, as the number of sleep bouts increases both during the day ($p = 0.002$ from the analysis of all deletion lines pooled across sexes compared to the wild type, Figure 5.5C, Supplementary Table S5.1H) and night ($p = 0.0001$ from the analysis of all deletion lines pooled across sexes compared to the wild type, Figure 5.5D, Supplementary Table S5.1H). Concomitant with increased day and night sleep, the *Uhg4* deletions on average have decreased length of activity bouts during the day ($p = 0.0061$, Figure 5.5E, Supplementary Table S5.1H as well as decreased total locomotor activity ($p = 0.0062$ from the analysis of all deletion lines pooled across sexes compared to the wild type, Figure 5.5F, Supplementary Table S5.1H).

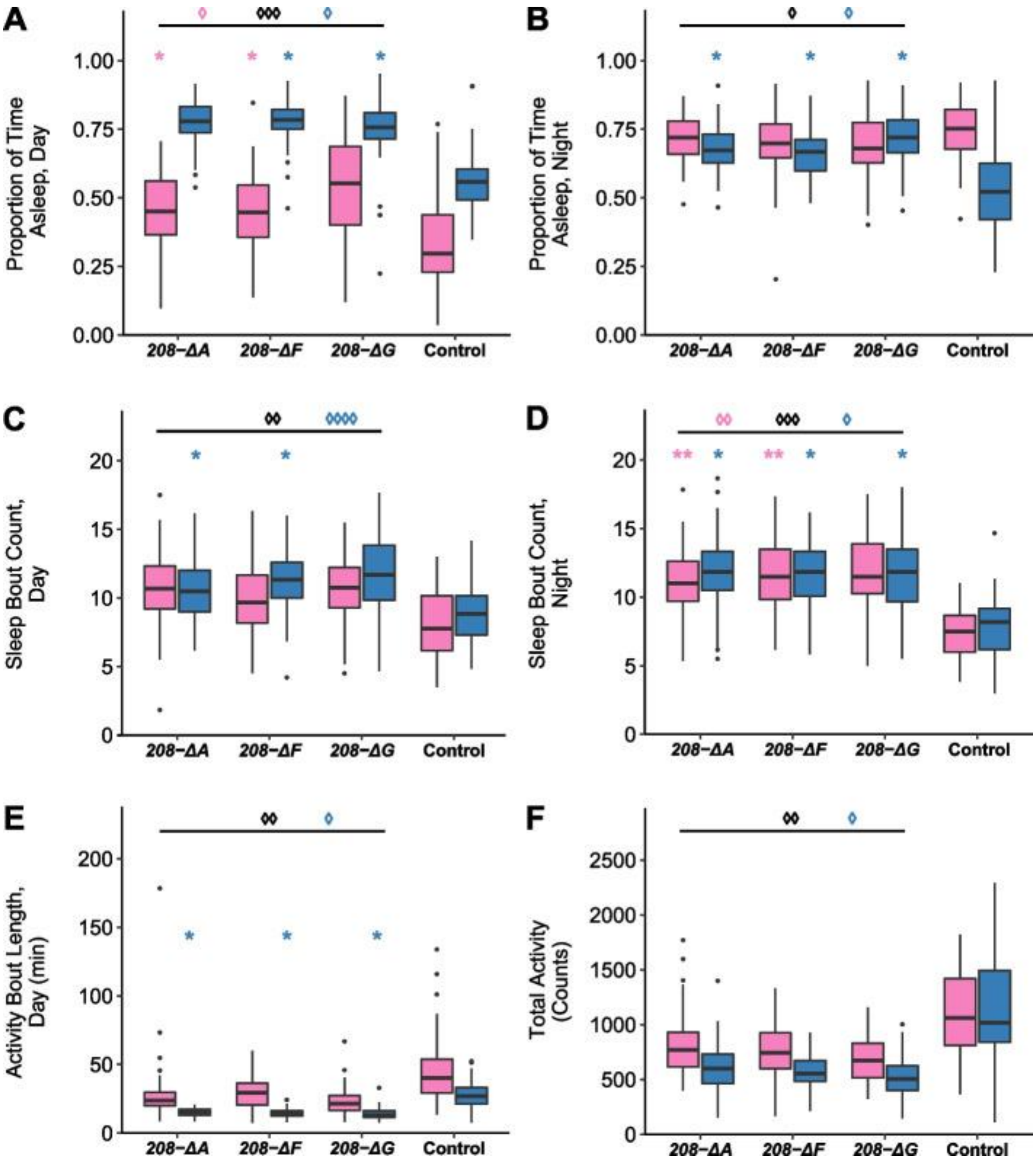


**Figure 5.5. Effects of *Uhg4* deletion on sleep and activity phenotypes. A** Boxplots showing the proportion of daytime sleep, **B** the proportion of nighttime sleep, **C** the number of sleep bouts at night, **D** the number of sleep bouts during the day, **E** total activity, and **F** activity bout length during the day. Blue boxes indicate males and pink

boxes indicate females. Day hours are from 6 am-6 pm. $N = 61–64$ flies per sex per line, except *208-ΔG*, which had only 32 females. See Supplementary Table S5.1 for ANOVAs. *p*-values on the figure indicated by an asterisk (*) represent comparisons of each sex to the control, by genotype. *p*-values on the figure indicated by a diamond (◊) represent comparisons of pooled *Uhg4* deletion genotypes versus the control. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

### *Effects of Uhg4 Deletions on Genome-Wide Gene Expression*

To gain insights into the mechanisms by which *Uhg4* exerts its pleiotropic effects, we assessed the consequences of deletion of *Uhg4* on genome-wide gene expression, and performed RNA-sequencing on *208-ΔA*, *208-ΔF*, *208-ΔG,* and DGRP_208 whole flies, separately for males and females. We performed factorial fixed effect ANOVAs for each of the 16,212 genes expressed in young adult flies that evaluate the significance of the main effects of the four *Uhg4* genotypes (Line), Sex, and the Line by Sex interaction. Plotting ordered raw *p*-values and adjusted *p*-values using a false discovery rate (FDR) correction using the Benjamini–Hochberg procedure (BH-FDR) against the number of tests revealed a non-monotonic relationship between raw *p*-values and adjusted *p*-values (Supplementary Figure S5.3a). This relationship caused an artificial inflation in the number of differentially expressed genes at BH-FDR < 0.1. Therefore, we used a BH-FDR thresholding approach to identify statistically significant genes at BH-FDR < 0.1. Briefly, after ordering the genes based on ascending raw *p*-values, we compared each gene's raw *p*-value to its BH-FDR critical value calculated as rank×(Q/number of tests) at both Q = 0.05 and Q = 0.1 (Supplementary Figure S5.3b). For both critical values, *p*-value thresholds were determined as the first occurrence of the raw *p*-value greater than critical values. Using this method, we identified 17 differentially expressed genes at a BH-FDR < 0.05 for the Line and/or Line by Sex terms (Supplementary Tables S5.2, S5.3). The top three differentially expressed genes were *Uhg4*, *snoRNA:Psi28S-2949,* and *snoRNA: Or-aca5*. The near-complete loss of *Uhg4* expression in the deletion genotypes is expected due to the deletions of the promoter region and exon 1. The two snoRNAs are located in the first two introns of *Uhg4*. Decreased expression of *snoRNA:Psi28S-2949* and *snoRNA:Or-aca5* suggests the *Uhg4* deletions affect regulatory sequences common to both snoRNAs, since their coding sequences are not altered (Figure 5.1). Ten of the 17 differentially expressed genes are computationally predicted genes and/or noncoding RNAs, including *Uhg4,* and have limited to no information on gene function. One of the 17 significantly differentially expressed genes was *insulin-like peptide 6* (*Ilp6*).

A total of 17 genes that are differentially expressed in *Uhg4* null genotypes is not sufficient to infer the function of *Uhg4* from the enrichment of Gene Ontologies (GO) and networks of the co-regulated genes. Therefore, we relaxed the significance threshold to BH-FDR < 0.1 for the Line and Line by Sex terms of the ANOVA models. This resulted in 180 differentially expressed genes (Supplementary Table S5.2). Notably, all genes significant for the Line by Sex terms also had significant Line effects (Supplementary Table S5.3). For these 180 genes, only one GO term (humoral immune response GO:0,006,959) was enriched at BH-FDR < 0.05 (Supplementary Table S5.4). However, 47 of the 180 genes did not map to a GO term. These analyses suggest that although deletion of *Uhg4* does change the transcriptome, many of these changes are associated with genes about which little, if anything, is known.

To fully quantify transcriptomic changes in the *Uhg4* deletion flies, we also assessed changes in expression of novel transcribed regions (NTRs) previously identified in the DGRP (including DGRP_208) [82]. Of the 18,581 NTRs analyzed, we found three (12) were differentially expressed at BH-FDR < 0.05 (0.1) for the Line term, and one NTR was differentially expressed for the Line by Sex term at BH-FDR < 0.1. Together with the protein-coding genes, we have a total 193 differentially expressed genes/NTRs in the *Uhg4* deletion lines (Supplementary Table S5.5).

We then performed k-means clustering to assess patterns of expression across genotypes, using k = 8, as it offered the largest number of clusters without redundancy of expression patterns across clusters. GO analysis of these k-means clusters emphasizes the sparsity of information currently known about the genes that are coregulated with *Uhg4*. Of the eight clusters, only four clusters (clusters 1, 3, 6, 7) have enriched GO terms (BH-FDR < 0.05). Cluster 1 is enriched for genes involved in immune response, cluster 3 is enriched for genes involved in response to stress and temperature, and clusters 6 and 7 are enriched for genes involved in cuticle structure (Supplementary Table S5.4). Cluster 1 consists of genes largely downregulated in deletion lines compared to the control, including *Uhg4,* as well as other noncoding RNAs and two NTRs (Supplementary Figure S5.4). Broadly, cluster 2 contains sexually dimorphic genes, whereas clusters 3 and 5 show genes upregulated in *208-ΔA* and *208-ΔG*, respectively, compared to other lines. *208-ΔF* displays intermediate expression patterns compared to *208-ΔA* and *208-ΔG* in clusters 3, 6, and 8 (Supplementary Figure S5.4).

In an effort to generate further hypotheses about the role of *Uhg4* and its co-regulated genes, we used the FlyBase database of known genetic and physical interactions within *D. melanogaster* [79] to generate interaction networks for the subset of 180 co-regulated genes at BH-FDR < 0.1. We also included first-degree interaction neighbors, genes, or proteins that are recorded in the database as directly interacting with at least one of the 180 focal genes. These networks revealed nine subclusters containing genes enriched for a broad spectrum of biological processes, including iron ion transport, fatty acid metabolism, temperature stress response, membrane trafficking, and morphogenesis (Figure 5.6, Supplementary Table S5.4). Cluster 9, which contains the genes *Ubx*, *dlp*, and *Pten*, among others, was enriched for hundreds of GO terms, far more than any other cluster, suggesting genes in this cluster are critical for a wide range of processes, including morphogenesis, cell differentiation, transcription factor signaling, sleep and activity, reproduction, stress response, and metabolism (Supplementary Table S5.4). Cluster 4 was not enriched for any GO terms and attempts at manual annotation did not reveal related functions for genes in this cluster (Supplementary Table S5.4). These results indicate that the lncRNA *Uhg4* contributes to diverse cellular functions.

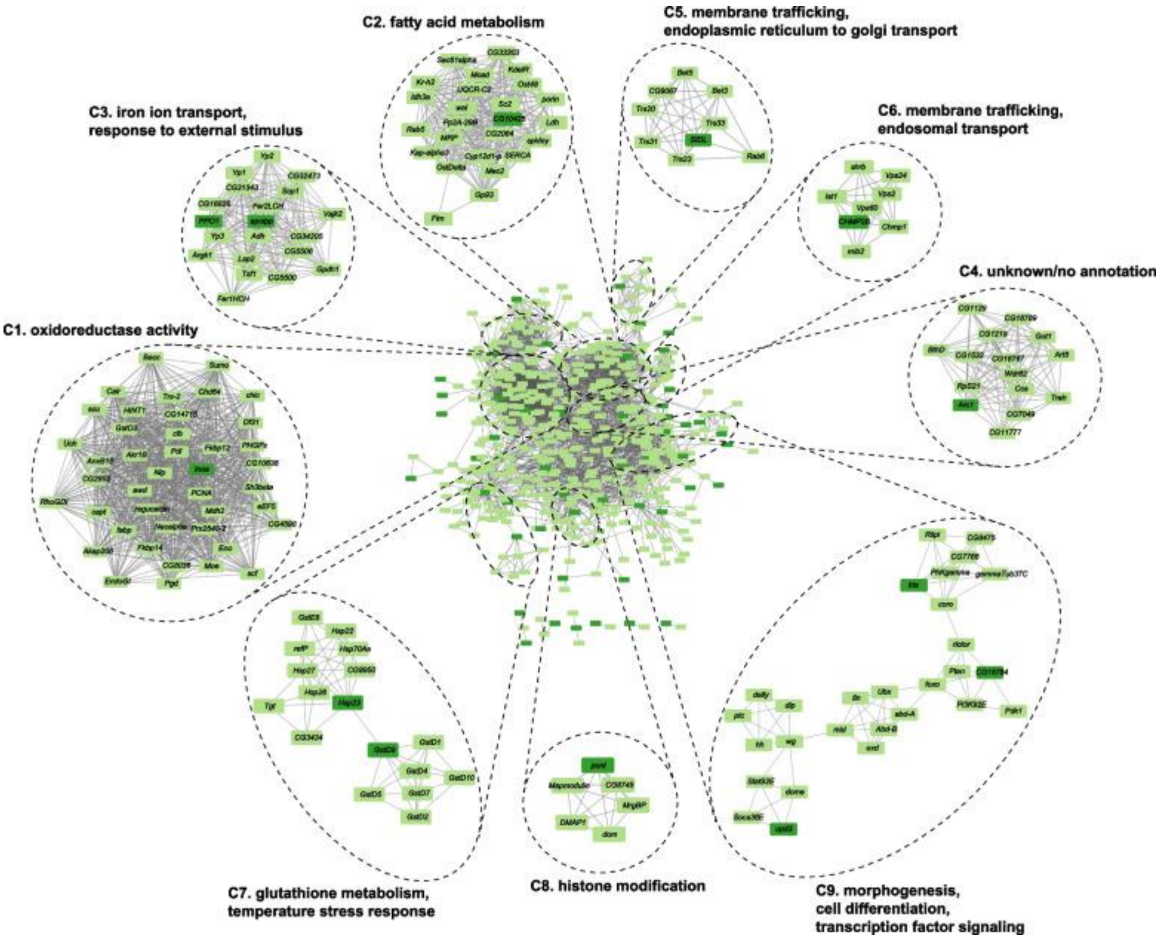

C2. fatty acid metabolism
C5. membrane trafficking,
endoplasmic reticulum to golgi transport
C3. iron ion transport,
response to external stimulus
C6. membrane trafficking,
endosomal transport
C4. unknown/no annotation
C1. oxidoreductase activity
C7. glutathione metabolism,
temperature stress response
C8. histone modification
C9. morphogenesis,
cell differentiation,
transcription factor signaling
Sec61alpha
CG33303
Kr-h2
UQCR-C2
Sc2
porin
Ldh
Rab5
CG2064
Kap-alpha3
Cyp12d1-p
SERCA
Gp93
Yp2
Yp1
CG32473
Yp3
Adh
CG34205
Argk1
Lsp2
Tsf1
Fer1HCH
Cat
Trx-2
HINT1
CG14715
Akr1B
Pdi
PCNA
Sh3beta
regucalcin
Mdh2
RhoGDI
fabp
Prx2540-2
Fkbp14
CG6036
Moe
Pgd
Bet5
CG9067
Bet3
Trs20
Trs33
Trs31
Trs23
Rab6
Vps24
Vps2
Vps60
CHMP2B
Chmp1
CG1129
CG18769
CG1219
CG18787
RpS21
CG7049
CG11777
GstE8
Hsp22
Hsp70Aa
Hsp27
CG8950
Hsp26
Hsp23
Tgt
CG3434
GstD1
GstD4
GstD10
GstD5
GstD7
GstD2
CG8745
DMAP1
dom
CG8475
CG7766
gammaTub37C
coro
rictor
Pten
foxo
Ubx
abd-A
Abd-B
PI3K92E
Pdk1
dally
dlp
wg
dome
Socs36E
upd3

**Figure 5.6. Interaction networks of genes coregulated with *Uhg4*.** Interaction networks (physical and genetic interactions) based on 180 genes with differential expression for the *Line* term (BH-FDR < 0.1). The full network (including interaction neighbors within 1 degree) is in the middle of the figure, with MCODE subclusters and subcluster GO enrichment annotations (Supplementary Table S5.4) on the perimeter. Dark green indicates the input 180 genes, and light green indicates an interaction neighbor. Names are Drosophila gene symbols. Numbers in parentheses indicate the cluster number.

## Discussion

We used *Uhg4*-knockout flies to assess the role of the lncRNA *Uhg4* across multiple fitness traits. We present evidence that *Uhg4* is critical for egg development and fertility and has pleiotropic effects on viability, development, stress responses, and sleep. Genome-wide gene expression data support the finding that *Uhg4* has pleiotropic effects, as *Uhg4* is coregulated with genes involved in a wide range of biological processes, including development, trafficking, metabolism, and stress response. We also observed that many of the genes differentially expressed upon loss of *Uhg4* are also noncoding RNAs in addition to predicted genes. We can speculate that *Uhg4* exerts its pleiotropic effects through broad-based gene regulatory networks. Further studies will be needed to obtain more detailed mechanistic insights into the effects on fitness traits of noncoding RNAs in *D. melanogaster*.

Although many of the coregulated genes themselves do not have associated GO terms, many biological processes and functions implicated by GO analysis align with observed changes in organismal phenotypes. Terms involving response to external stimuli such as temperature and ethanol, as well as terms relating to immune system response, are enriched in multiple k-means and interaction network clusters. These terms align with the increased susceptibility of the *Uhg4* deletion flies to extreme hot and cold temperatures, as well as the decreased susceptibility to ethanol-mediated stressors. Interestingly, *Alcohol dehydrogenase* (*Adh*) interacts with at least two genes coregulated with *Uhg4,* providing a possible mechanistic link between *Uhg4* and ethanol response; *Uhg4* has been previously implicated in developmental ethanol exposure, which also results in decreased susceptibility to ethanol exposure in adult flies [80]. Although not annotated in Figure 5.6, interaction cluster nine is enriched for genes involved in wing morphogenesis and gamete generation (Supplementary Table S5.4), which could explain the changes to wing position and sterility phenotypes, respectively, observed in *Uhg4* deletion flies. We constructed interaction networks from an input of 20 differentially expressed genes and NTRs (BH-FDR $< 0.05$) including neighbors within at least 1 degree (Supplementary Figure S5.5A) and less stringent networks with input of 180 coregulated genes (BH-FDR $< 0.1$) including neighbors within at least 2 degrees (Supplementary Figure S5.5B). Cluster 9 in Supplementary Table S5.4, and its more stringent 2-degree interaction counterpart (Cluster 3 in Supplementary Figure S5.5B), are enriched for hundreds of GO terms (Supplementary Table S5.4), indicating that these genes have wide-ranging impacts. Thus, the effects of *Uhg4* deletion may extend to additional traits that we did not assess, such as iron ion transport, or other intracellular phenotypes.

Based on the DAPI-stained ovary images (Figure 5.2) showing a lack of late-stage oocytes, we hypothesize that *Uhg4*-deletion females are capable of laying eggs but do not develop late-stage oocytes. This oocyte development phenotype is supported by the high expression of *Uhg4* in ovaries. However, this does not explain why embryos from wild-type females and *Uhg4* deletion males fail to develop.

*Bootlegger* (*Boot*), a gene located immediately upstream of and in opposite orientation to *Uhg4*, is also critical for proper egg development [83, 84]. Our Sanger sequencing data indicate that *Boot* is intact in all deletions and our RNAseq data do not show *Boot* to be differentially expressed. Furthermore, unlike *Uhg4*, *Boot* is minimally expressed in males and would unlikely be responsible for the sterility observed in *Uhg4*-deficient males. We are therefore confident that the phenotypes observed in our *Uhg4* deletion flies can be attributed to *Uhg4*, though it is possible that *Uhg4* and *Boot* share promoter elements in their intergenic region.

The role of some snoRNA host genes such as *Uhg4* was thought to facilitate transcription and splicing of snoRNAs [85–87]. Based on our results, we hypothesize that *Uhg4* has roles independent of hosting snoRNAs, as most of its snoRNAs are not differentially expressed in *Uhg4* deletion flies. The abundance of genes/NTRs coregulated with *Uhg4* that are noncoding RNAs and/or have no known function makes speculation about the possible functional mechanisms by which *Uhg4* affects a wide range of pleiotropic phenotypes challenging. *Uhg4* could bind directly to DNA or transcription factors to modulate transcriptional regulation, acting in *cis* to regulate the expression of *snoRNA:Psi28S-2949* and *snoRNA:Or-aca5*, or in *trans* to regulate the expression of other genes we observed to be differentially expressed. *Uhg4* may also serve an oncogenic role, as overexpression of *Uhg4* in a Drosophila cell line is associated with tumor growth, and*Uhg4* is a downstream target of *Myc* [88]. *Uhg4* may also act to regulate gene expression at an epigenetic level via histone modifications (Cluster 8, Figure 5.6). Other lncRNAs in Drosophila that are important for thermotolerance are essential for remobilization of heterogenous ribonucleoprotein particles (hnRNPs) [41, 44]. *Uhg4*, as a host gene for snoRNAs, which also form ribonucleoproteins, might also modulate response to thermal stressors via ribonucleoproteins. *Uhg4* could also act in a similar manner to the transcript of *oskar*, which is critical for oogenesis in Drosophila [54, 89, 90]. *oskar* RNA facilitates oogenesis through multiple mechanisms, as it binds a translational regulator at one locus, has a separate 3' region critical for proper egg-laying, and may also be involved in scaffolding of ribonucleoproteins [89].

*Uhg4* is a unique example of a lncRNA with pleiotropic functions that is indispensable for viability and reproduction in *D. melanogaster*.

## Methods

### *Generation of Uhg4 null alleles*

We used the flyCRISPR target finder [91] to design gRNAs flanking *Uhg4* Exon 1 (upstream: 5'- GAAGTAAAACTTCTTTGCACTGG -3'; downstream: 5'- GTAAGTATTATAGATATGATAGG -3') that did not overlap with known genes and did not have predicted off-target effects, resulting in a 685 bp deletion. We then synthesized a single guide RNA (gRNA) vector containing both gRNAs [92]. Briefly, synthesized phosphorylated gRNAs with complementary sequences were ligated to form double-stranded gRNAs with CTTC overhangs. One double-stranded gRNA was cloned into *pBFv-U6.2* (Addgene #138,400), and the other double-stranded gRNA was cloned into *pBFv-U6.2B* (Addgene #138,401). Using flanking *Eco*RI and *Not*I sites, the resulting *U6* promoter and gRNA within *pBFV-U6.2* were excised and ligated into the *U6.2B* vector, creating a double-gRNA vector. Sanger sequencing confirmed the proper insertion of gRNAs.

The completed dual gRNA vector and *pBFv-nosP-Cas9* (Addgene #138,402) were purified and co-injected (BestGene Inc., Chino Hills, CA) into at least 300 embryos from two different *D. melanogaster* Genetic Reference Panel (DGRP) lines [81, 93], DGRP_208 and DGRP_705. These lines have minimal heterozygosity and are homosequential for the standard karyotype for all common inversions; DGRP_208 is also free of *Wolbachia* infection [81].

To preserve the inbred genetic background of the DGRP lines, we screened flies for the presence of a deletion by individually isolating DNA from clipped wings of virgin female and male flies [94]. Fly wings were covered with 10 µL of 400 µg/mL protease K in a 10 mM Tris–Cl at pH 8.2, 1 mM EDTA and 25 mM NaCl buffer and incubated at 37 °C for 2 h followed by 95 °C for 2 min. We used 2µL of the resulting DNA mixture in a PCR reaction with primers (Left: 5’- CTAGCACGGAACCCTGGAAAT -3’; Right: 5’- GCAGCGCCTAGTAATCACAGA -3’) according to ApexRedTaq (Genesee Scientific, El Cajon, CA) manufacturer instructions, with a 61 °C annealing temperature for 30 s and

30 cycles. The deletion mutations *Uhg4*$^{208-\Delta A}$, *Uhg4*$^{208-\Delta F}$, *Uhg4*$^{208-\Delta G}$, and *Uhg4*$^{705-\Delta B}$ (hereafter referred to as *208-ΔA*, *208-ΔF*, *208-ΔG* and *705-ΔB*, respectively) were isolated and placed over a *CyO* balancer chromosome in the appropriate genetic background. *Uhg4* wild-type lines (DGRP_208 and DGRP_705) are used as controls in this study.

### *Drosophila culture*

Flies were reared at 25 °C with 50% humidity on standard cornmeal-molasses-agar medium (Genesee Scientific, El Cajon, CA), supplemented with yeast. Flies were maintained at controlled density on a 12-h light–dark cycle (lights on at 6 am). Unless otherwise indicated, all behavioral assays were performed on 3–5 day old homozygous flies from 8:30 am to 11:30 am.

### *Sanger sequencing*

DNA from homozygous mutant flies was sequenced by the Sanger chain termination method using a BigDye Terminator Kit v3.1 (ThermoFisher Scientific, Waltham, MA), with the same primers used to identify deletions. Sequencing was performed on an Applied Biosystems 3730xl DNA Analyzer (ThermoFisher Scientific, Waltham, MA).

### *Quantitative real-time PCR*

Flies were flash-frozen for RNA extraction at 3–5 days of age. Each sample contained 30 whole flies and was homogenized using a Fisherbrand™ Bead Mill (ThermoFisher

Scientific, Waltham MA). RNA was extracted using the Direct-zol RNA Miniprep kit (Zymo Research, Irvine, CA), resuspended in RNase-free water, and kept at -80 °C until further use. cDNA was synthesized using the iScript™ Reverse Transcription Supermix for RT-qPCR (Bio-Rad Laboratories, Inc., Hercules, CA) according to the manufacturer's instructions. Quantitative real-time PCR to detect expression was performed on three biological replicates and two technical replicates per sample (except *208-ΔA*, which had two biological replicates), using SYBR™ Green PCR Master Mix (ThermoFisher Scientific, Waltham, MA) and primers (*Uhg4*-Forward: 5'-TCGGTCTTTCGATTTGGATT -3'; *Uhg4*-Reverse: 5'-TGTGTTAGTGAGCCACGTTTG -3', spanning exons 4–5 of *Uhg4*; GAPDH-Left: 5'-CTTCTTCAGCGACACCCATT -3'; GAPDH-Right: 5'-ACCGAACTCGTTGTCGTACC -3') according to the manufacturer's instructions. Percent knockdown was assessed using *ΔΔct* [95]. No amplification was observed in non-template negative controls.

*Viability and development time*

We placed 75 male–female pairs of flies from each of the DGRP_208 *Uhg4* genotypes (DGRP_208/*CyO*; *208-ΔA*/*CyO*, *208-ΔF*/*CyO*, *208-ΔG*/*CyO*) in large embryo collection cages (Genesee Scientific, San Diego, CA), supplemented with fresh yeast paste and grape juice-agar plates containing 3% agar and 1.2% sucrose in a 25% Welch's Grape Juice concentrate solution every 12 h. After 36 h, we placed 50 0–12 h embryos in each of 25 vials per genotype. From days 9–13 after egg-laying, we collected adult flies once daily and recorded the sex, wing phenotype, and day of emergence for each fly, including flies that eclosed, but died in the food. We calculated relative egg-adult viability ($v$) for each vial as the proportion of homozygous *Uhg4* adults (1-$r$) relative to the proportion of *Uhg4*/*CyO* heterozygote adults ($r$) as $v = 2(1\text{-}r)/r$ [96].

### *Fertility*

We assessed the fertility of homozygous males and females from the DGRP_208 *Uhg4* deletion lines by crossing virgin deletion mutant flies to virgin DGRP_208 flies, respectively, of the opposite sex. In addition, we also performed crosses of males and females within each *Uhg4* deletion line. We set up four vials each with four males and four virgin females for each genotype. For each cross, qualitative observations (presence/absence) were made for each stage of development (embryos, first instar larvae, second instar larvae, third instar larvae, pupae, adult flies). We did not perform a formal fertility assay with the *705-ΔB* deletion line due to the very low viability of *705-ΔB* homozygotes.

### *Mating behavior*

For each DGRP_208 *Uhg4* genotype (*208-ΔA*, *208-ΔF*, *208-ΔG*, DGRP_208), we placed 22–24 pairs of virgin flies in separate mating chambers [97, 98] to acclimate overnight. Between 8 and 10 am, fly pairs were united and video-recorded for 30 min. Copulation duration and mating latency were recorded in seconds for each fly pair. We only included flies that mated in the analyses of mating behaviors and recorded the number of pairs that did not mate during the 30-min testing period. We selected *208-ΔG* as a representative *Uhg4* mutant line and assessed the mating behaviors of crosses of *208-ΔG* males to DGRP_208 virgin females and *208-ΔG* virgin females to DGRP_208 males.

### *Survival after heat shock*

Vials with 2 ml of medium and 9 homozygous flies per sex and *Uhg4* genotype were placed in a 37 °C incubator for 2 or 3 h beginning at 1 pm, with five replicate vials per sex per genotype per exposure time. After the heat shock, flies were transferred to fresh vials with 2 ml of medium and allowed to recover overnight. Twenty-four hours after the heat shock, the number of surviving flies in each vial was recorded.

### *Chill coma recovery time*

For each genotype we placed four vials containing 15 flies without medium on ice for 3 h, sexes separately, and allowed the flies to recover in a 6-well cell culture plate (five flies per well, two genotypes per plate) for 30 min. We recorded the time until each fly righted itself by standing up [99]. Only flies that recovered within 30 min of being removed from the ice were included in the analysis ($n = 48$–57 flies per sex per line), although we also recorded the number of flies that did not recover.

### *Ethanol sensitivity*

We assessed the time to sedation in response to acute ethanol exposure [100] on 44–52 flies per sex per *Uhg4* genotype. Briefly, flies were aspirated into a 24-well cell culture plate and placed opposite a 24-well cell culture plate containing 100% ethanol, separated by a layer of fine screen mesh. We recorded the time to sedation, defined as the moment each fly loses postural control.

### *Sleep and locomotor activity*

We collected sleep and locomotor activity data using Drosophila Activity Monitors (DAM) (TriKinetics, Waltham, MA). Briefly, we placed 1–2 day-old flies into DAM monitor tubes containing 2% agar with 5% sucrose, with two DAM monitors per sex per *Uhg4* genotype, for a total of 64 flies per sex and genotype. Sleep and activity data were recorded on days 3–8 of the fly lifespan on a 12-h light–dark cycle. Sleep was defined as at least 5 min of inactivity. Only flies that survived the entire testing period were included in analyses, resulting in 61–64 flies per sex per *Uhg4* genotype (except for *208-ΔG* females, for which there was only one replicate of 32 individuals). The DAM data were initially processed with Shiny-RDAM [101] and resulting raw sleep and activity output files were downloaded for further statistical analysis.

### *Statistical analyses*

Unless otherwise indicated, all behavioral assays were analyzed using the “PROC MIXED” command (for a mixed-effects model) or “PROC GLM” command (for a pure fixed-effects model) in SAS v3.8 as a Type III Analysis of Variance (ANOVA). Where appropriate, flies were randomized to avoid positional or time-related effects.

Development time (day of eclosion) was analyzed according to the model $Y = \mu + L + S + G + L \times S + L \times G + S \times G + L \times S \times G + Rep(L) + S \times Rep(L) + G \times Rep(L) + S \times G \times Rep(L) + \varepsilon$. Heat shock (percent of flies that survived) and all sleep and activity phenotypes (e.g. locomotor activity, bout count, percent of time asleep) were analyzed according to the model $Y = \mu + L + S + L \times S + Rep(L \times S) + \varepsilon$. Chill coma recovery time (time to recover, in seconds) and ethanol sensitivity (time to sedation, in seconds) were analyzed according to the model $Y = \mu + L + S + L \times S + \varepsilon$. Mating latency (in seconds), mating duration (in seconds), and viability (viability coefficient $v$) were assessed according to the ANOVA model $Y = \mu + L + \varepsilon$. In these models, $Y$ is the phenotype of interest, $\mu$ is the overall mean, $L$ is the fixed effect of Line (DGRP_208, *208-ΔA*, *208-ΔF*, *208-ΔG*), $S$ is the fixed effect of sex (male, female), $G$ is the fixed effect of balancer genotype (*Uhg4*/*CyO* heterozygote, *Uhg4* homozygote), *Rep* is replicate nested within lines, and $\varepsilon$ is the residual error. We performed pairwise comparisons of each *Uhg4* deletion line (*208-ΔA*, *208-ΔF*, or *208-ΔG*) with the *Uhg4* wild type control, as well as a pooled comparison across all deletion lines compared to the control. For mating phenotypes, we also compared *208-ΔG* females x DGRP_208 males and DGRP_208 females × *208-ΔG* males to the control. We used Fisher's exact tests to assess the proportion of flies that mated and the proportion of flies that recovered from a chill-induced coma in mating and chill coma recovery time, respectively using the R *stats* package. Models were also run separately by sex. For development time, models were also run separately for *Uhg4*/*CyO* heterozygotes and *Uhg4* homozygotes.

*Ovary dissection*

We placed mated females from representative *Uhg4* deletion lines and their controls (*208-ΔF*, *208-ΔG*, DGRP_208) in fresh food vials supplemented with yeast paste every 12 h for 36 h prior to dissection. Flies were dissected in 1X PBS and ovarioles were gently separated. Ovaries were fixed in 4% paraformaldehyde for 15 min, followed by three 15-min washes in PBS with 0.2% Triton X-100. Following a final 15-min wash in PBS, ovaries were stained with DAPI (Invitrogen) (1 μg/mL) for 10 min and mounted with ProLong Gold (Invitrogen) immediately after the final PBS wash. Ovaries were imaged on an Olympus Fluoview FV3000 microscope at 20 × magnification. Images were processed in Fiji [102].

*RNA sequencing*

We prepared libraries for RNA sequencing from each RNA sample used in the RT-qPCR analyses according to Universal RNA-Seq with NuQuant + UDI (Tecan Genomics, Inc., Redwood City, CA) manufacturer instructions. Specifically, 100 ng of total RNA was converted into cDNA via integrated DNase treatment. Second strand cDNA was fragmented using a Covaris ME220 Focused-ultrasonicator (Covaris, Woburn, MA) to 350 bp. Drosophila AnyDeplete probes were used to deplete remaining ribosomal RNA and final libraries were amplified using 17 PCR cycles. We used TapeStation High Sensitivity D1000 Screentape (Agilent Technologies, Inc., Santa Clara, CA) and a Qubit™ 1X dsDNA HS Assay kit (Invitrogen, Carlsbad, CA) to measure the final library insert sizes and concentration, respectively. Final libraries were diluted to 4 nM and sequenced on a NovaSeq6000 (Illumina, Inc., San Diego, CA) using an S1 flow cell. We sequenced two biological replicates per sex per line (*208-ΔA*, *208-ΔF*, *208-ΔG*, DGRP_208), with ~ 20–74 million reads per sample.

Barcoded reads were demultiplexed using the NovaSeq Illumina BaseSpace sequencing pipeline and merged across S1 flow cell lanes. We used the *AfterQC* pipeline (version 0.9.7) [103] to filter out low-quality, short, and adapter reads and the bbduk command within the BBTools package [104] to detect levels of rRNA contamination. We used GMAP-GSNAP [105] to align filtered reads to the *D. melanogaster* reference genome v6.13 and the featurecounts pipeline from the Subread package [106] to count unique alignments to Drosophila genes. Expression for novel transcribed regions (NTRs) was estimated by first compiling a list of NTRs detected from RNA sequencing of young adult DGRP flies [82]. The coordinates of these NTRs were converted from R5 to R6 using the Coordinate Converter tool on FlyBase. A new gene transfer format file was constructed using the coordinate-converted NTR gene models and was used in conjunction with the alignment files for expression estimation. Counts data for each sample were filtered to omit genes for which the median count was less than two, as well as genes for which the proportion of null values across all samples was less than 0.25. The data were then normalized for gene length and library size using Ge-TMM [107]. Filtered and normalized data were analyzed for differential expression using the “PROC glm” command in SAS v3.8 (Cary, NC) according to the ANOVA model $Y = \mu + L + S + L \times S + \varepsilon$, where Line ($L$), Sex ($S$), and Line × Sex ($L \times S$) are fixed effects,

$Y$ is gene expression, $\mu$ is the overall mean, and $\varepsilon$ is residual error. A false discovery rate (FDR) correction using the Benjamini–Hochberg procedure (BH-FDR) for multiple tests was applied across all genes to determine the subset of differentially expressed genes significant at BH-FDR < 0.05 and BH-FDR < 0.1 for either the Line or Line x Sex terms. NTR expression counts were analyzed using the same approach described above but as a separate dataset. For the bulk RNA sequencing data, plotting ordered raw *p*-values and BH-FDR adjusted *p*-values against the number of tests revealed a non-monotonic relationship between raw *p*-values and adjusted *p*-values. Therefore, we used a BH-FDR thresholding approach to identify genes with statistically significant *p*-values from the ANOVA model. Briefly, after ordering the genes based on ascending raw *p*-values, we compared each gene's raw *p*-value to FDR critical values calculated as rank*Q/number of tests at both Q = 0.05 and Q = 0.1. For both critical values, *p*-value thresholds were determined as the first occurrence of the raw *p*-values greater than critical values. Genes with raw *p*-values below the *p*-value threshold at critical values Q = 0.05 and Q = 0.1 were considered for downstream analyses. We used the resulting 180 genes significant at FDR < 0.1 for network construction and included the 13 differentially expressed NTRs (BH-FDR < 0.1) with these 180 genes for k-means clustering.

We performed k-means clustering (k = 8, average linkage algorithm) on the Ge-TMM-normalized least squares means of the 193 coregulated genes. We performed iterative k-means clustering with different k values to determine the largest number of clusters without redundant expression patterns across clusters. We used Cytoscape v3.9.1 and the interaction networks from FlyBase [79] (FB_2021_05) to create protein and genetic interaction networks including first-degree neighbors, clustered via MCODE score [108] where applicable. Cluster annotations are based on significantly enriched Gene Ontology (GO) terms. We performed GO statistical overrepresentation analyses with GO: Biological Process Complete, Molecular Function Complete, and Reactome Pathway terms (GO Ontology database released 2021–11-16) using Panther db v16.0 [109] using Fisher Exact tests with BH-FDR correction.

*Data Availability Statement*

All high throughput sequencing data are deposited in GEO GSE199865.

Sanger sequencing, qPCR, and raw behavioral data are available on GitHub at rebeccamacpherson/Pleitropic_fitness_effects_Uhg4_rawdata

## Acknowledgements

We thank Miller Barksdale and Patrick Freymuth for their technical assistance, Dr. Sneha Mokashi for experimental advice and Jonathon Thomalla and Drs. Mariana Wolfner, Melissa White, and Deeptiman Chatterjee for their insights into ovary biology. Additionally, we would like to thank the Greenwood Genetic Center and Dr. Heather Flanagan-Steet for the use of their Sanger sequencing and microscope facilities.

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM139769 and Award Number R01GM128974 to RRHA and TFCM and by the National Institute of Child Health and Development under Award Number F31HD106719 to RAM. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

## References

1. Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the *H19* gene may function as an RNA. *Mol Cell Biol.* 1990;10: 28-36. doi: 10.1128/mcb.10.1.28-36.1990.

2. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, et al. The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell.* 1992;71: 515-26. doi: 10.1016/0092-8674(92)90519-I.

3. Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, et al. The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*. 1992;71: 527-42. doi: 10.1016/0092-8674(92)90520-M.

4. Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. Genetics. 2013;193: 651-69. doi: 10.1534/genetics.112.146704.

5. McHugh CA, Chen C, Chow A, Surka CF, Tran C, McDonel P, et al. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*. 2015;521: 232-36. doi: 10.1038/nature14443.

6. Rom A, Melamed L, Gil N, Goldrich MJ, Kadir R, Golan M, et al. Regulation of *CHD2* expression by the *Chaserr* long noncoding RNA gene is essential for viability. *Nat Commun*. 2019;10: 5092. doi: 10.1038/s41467-019-13075-8.

7. Pandya-Jones A, Markaki Y, Serizay J, Chitiashvili T, Mancia Leon WR, Damianov A, et al. A protein assembly mediates *Xist* localization and gene silencing. *Nature*. 2020;587: 145-51. doi: 10.1038/s41586-020-2703-0.

8. Statello L, Guo C, Chen L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol.* 2021;22: 96-118. doi: 10.1038/s41580-020-00315-9.

9. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, et al. The *Xist* lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science*. 2013;341: 1237973. doi: 10.1126/science.1237973.

10. Li X, Zhou B, Chen L, Gou L, Li H, Fu X. GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol*. 2017;35: 940-50. doi: 10.1038/nbt.3968.

11. Mumbach MR, Granja JM, Flynn RA, Roake CM, Satpathy AT, Rubin AJ, et al. HiChIRP reveals RNA-associated chromosome conformation. *Nat Methods.* 2019;16: 489-92. doi: 10.1038/s41592-019-0407-x.

12. Sánchez Y, Segura V, Marín-Béjar O, Athie A, Marchese FP, González J, et al. Genome-wide analysis of the human *p53* transcriptional network unveils a lncRNA tumour suppressor signature. *Nat Commun.* 2014;5: 5812. doi: 10.1038/ncomms6812.

13. Michelini F, Pitchiaya S, Vitelli V, Sharma S, Gioia U, Pessina F, et al. Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. *Nat Cell Biol.* 2017;19: 1400-11. doi: 10.1038/ncb3643.

14. Uroda T, Anastasakou E, Rossi A, Teulon J, Pellequer J, Annibale P, et al. Conserved Pseudoknots in lncRNA *MEG3* Are Essential for Stimulation of the p53 Pathway. *Mol Cell*. 2019;75: 982-95.e9. doi: 10.1016/j.molcel.2019.07.025.

15. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol Cell*. 2009;33: 717-26. doi: 10.1016/j.molcel.2009.01.026.

16. Hacisuleyman E, Shukla CJ, Weiner CL, Rinn JL. Function and evolution of local repeats in the *Firre* locus. *Nat Commun.* 2016;7: 11021. doi: 10.1038/ncomms11021.

17. Wu H, Yin Q, Luo Z, Yao R, Zheng C, Zhang J, et al. Unusual Processing Generates SPA LncRNAs that Sequester Multiple RNA Binding Proteins. *Mol Cell.* 2016;64: 534-48. doi: 10.1016/j.molcel.2016.10.007.

18. Fei J, Jadaliha M, Harmon TS, Li ITS, Hua B, Hao Q, et al. Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. J *Cell Sci.* 2017;130: 4180-92. doi: 10.1242/jcs.206854.

19. Soshnev AA, Ishimoto H, McAllister BF, Li X, Wehling MD, Kitamoto T, et al. A Conserved Long Noncoding RNA Affects Sleep Behavior in Drosophila. *Genetics.* 2011;189: 455-68. doi: 10.1534/genetics.111.131706.

20. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet.* 2016;17: 272-83. doi: 10.1038/nrg.2016.20.

21. Hu WL, Jin L, Xu A, Wang YF, Thorne RF, Zhang XD, et al. GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. *Nat Cell Biol.* 2018;20: 492-502. doi: 10.1038/s41556-018-0066-7.

22. Yin Q, Yang L, Zhang Y, Xiang J, Wu Y, Carmichael G, et al. Long Noncoding RNAs with snoRNA Ends. *Mol Cell.* 2012;48: 219-30. doi: 10.1016/j.molcel.2012.07.033.

23. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*. 2016;539: 452-5. doi: 10.1038/nature20149.

24. Yap K, Mukhina S, Zhang G, Tan JSC, Ong HS, Makeyev EV. A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. *Mol Cell*. 2018;72: 525-40.e13. doi: 10.1016/j.molcel.2018.08.041.

25. Pisignano G, Ladomery M. Epigenetic Regulation of Alternative Splicing: How LncRNAs Tailor the Message. *Noncoding RNA.* 2021;7: 21. doi: 10.3390/ncrna7010021.

26. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol.* 2007;9: 660-5. doi: 10.1038/ncb1595.

27. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP. Peptides Encoded by Short ORFs Control Development and Define a New Eukaryotic Gene Family. *PLoS Biol.* 2007;5: e106. doi: 10.1371/journal.pbio.0050106.

28. Hartford CCR, Lal A. When Long Noncoding Becomes Protein Coding. *Mol Cell Biol*. 2020;40. doi: 10.1128/MCB.00528-19.

29. Noh JH, Kim KM, Abdelmohsen K, Yoon J, Panda AC, Munk R, et al. HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. *Genes Dev*. 2016;30: 1224-39. doi: 10.1101/gad.276022.115.

30. Krause HM. New and Prospective Roles for lncRNAs in Organelle Formation and Function. *Trends Genet*. 2018;34: 736-45. doi: 0.1016/j.tig.2018.06.005.

31. Lubelsky Y, Ulitsky I. Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature*. 2018;555: 107-11. doi: 10.1038/nature25757.

32. Azam S, Hou S, Zhu B, Wang W, Hao T, Bu X, et al. Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. *RNA Biol.* 2019;16: 1001-9. doi: 10.1080/15476286.2019.1620061.

33. Yan P, Luo S, Lu JY, Shen X. Cis- and trans-acting lncRNAs in pluripotency and reprogramming. *Curr Opin Genet Dev*. 2017;46: 170-8. doi: 10.1016/j.gde.2017.07.009.

34. Ulitsky I, Shkumatava A, Jan C, Sive H, Bartel D. Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell*. 2012;151: 684-6. doi: 10.1016/j.cell.2012.10.002.

35. Kirk JM, Kim SO, Inoue K, Smola MJ, Lee DM, Schertzer MD, et al. Functional classification of long non-coding RNAs by k-mer content. *Nat Genet.* 2018;50: 1474-82. doi: 10.1038/s41588-018-0207-8.

36. Guo C, Ma X, Xing Y, Zheng C, Xu Y, Shan L, et al. Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell*. 2020;181: 621-36.e22. doi: 10.1016/j.cell.2020.03.006.

37. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov*. 2011;1: 391-407. doi: 10.1158/2159-8290.CD-11-0209.

38. Peng W, Koirala P, Mo Y. LncRNA-mediated regulation of cell signaling in cancer. *Oncogene*. 2017;36: 5661-7. doi: 10.1038/onc.2017.184.

39. Carlevaro-Fita J, Lanzós A, Feuerbach L, Hong C, Mas-Ponte D, Pedersen JS, et al. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun Biol*. 2020;3: 56. doi: 10.1038/s42003-019-0741-7.

40. Yang M, Lu H, Liu J, Wu S, Kim P, Zhou X. lncRNAfunc: a knowledgebase of lncRNA function in human cancer. *Nucleic Acids Res*. 2022;50: D1295-D1306. doi: 10.1093/nar/gkab1035.

41. Lo Piccolo L, Yamaguchi M. RNAi of arcRNA *hsrω* affects sub-cellular localization of Drosophila FUS to drive neurodiseases. *Exp Neurol*. 2017;292: 125-34. doi: 10.1016/j.expneurol.2017.03.011.

42. Muraoka Y, Nakamura A, Tanaka R, Suda K, Azuma Y, Kushimura Y, et al. Genetic screening of the genes interacting with Drosophila FIG4 identified a novel link between CMT-causing gene and long noncoding RNAs. *Exp Neurol.* 2018;310: 1-13. doi: 10.1016/j.expneurol.2018.08.009.

43. Wei C, Luo T, Zou S, Wu A. The Role of Long Noncoding RNAs in Central Nervous System and Neurodegenerative Diseases. *Front Behav Neurosci*. 2018;12: 175. doi: 10.3389/fnbeh.2018.00175.

44. Lakhotia SC, Mallik M, Singh AK, Ray M. The large noncoding *hsrω-n* transcripts are essential for thermotolerance and remobilization of hnRNPs, HP1 and RNA polymerase II during recovery from heat shock in Drosophila. *Chromosoma*. 2012;121: 49-70. doi: 10.1007/s00412-011-0341-x.

45. Kukharsky MS, Ninkina NN, An H, Telezhkin V, Wei W, Meritens CRd, et al. Long non-coding RNA Neat1 regulates adaptive behavioural response to stress in mice. *Transl Psychiatry*. 2020;10: 171. doi: 10.1038/s41398-020-0854-2.

46. Gao J, Chow EWL, Wang H, Xu X, Cai C, Song Y, et al. LncRNA DINOR is a virulence factor and global regulator of stress responses in *Candida auris*. *Nat Microbiol*. 2021;6: 842-851.

47. Zeni PF, Mraz M. LncRNAs in adaptive immunity: role in physiological and pathological conditions. *RNA Biol*. 2021;18: 619-32. doi: 10.1080/15476286.2020.1838783.

48. Nakamura A, Amikura R, Mukai M, Kobayashi S, Lasko PF. Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. *Science.*1996;274: 2075-9. doi: 10.1126/science.274.5295.2075.

49. Hardiman KE, Brewster R, Khan SM, Deo M, Bodmer R. The *bereft* Gene, a Potential Target of the Neural Selector Gene *cut*, Contributes to Bristle Morphogenesis. *Genetics*. 2002;161: 231-47.

50. Pease B, Borges AC, Bender W. Noncoding RNAs of the Ultrabithorax Domain of the Drosophila Bithorax Complex. *Genetics*. 2013;195: 1253-64. doi: 10.1534/genetics.113.155036.

51. Chen B, Zhang Y, Zhang X, Jia S, Chen S, Kang L. Genome-wide identification and developmental expression profiling of long noncoding RNAs during Drosophila metamorphosis. *Sci Rep*. 2016;6: 23330.

52. Kim M, Faucillion M, Larsson J. *RNA-on-X 1* and *2* in *Drosophila melanogaster* fulfill separate functions in dosage compensation. *PLoS Genet*. 2018;14: e1007842. doi: 10.1371/journal.pgen.1007842.

53. Xu M, Xiang Y, Liu X, Bai B, Chen R, Liu L, et al. Long noncoding RNA SMRG regulates Drosophila macrochaetes by antagonizing *scute* through E(spl)mβ. *RNA Biol*. 2019;16: 42-53. doi: 10.1080/15476286.2018.1556148.

54. Jenny A, Hachet O, Závorszky P, Cyrklaff A, Weston MD, Johnston DS, et al. A translation-independent role of *oskar* RNA in early Drosophila oogenesis. *Development*. 2006;133: 2827-33. doi: 133/15/2827 [pii].

55. Wen K, Yang L, Xiong T, Di C, Ma D, Wu M, et al. Critical roles of long noncoding RNAs in Drosophila spermatogenesis. *Genome Res*. 2016;26: 1233-44. doi: 10.1101/gr.199547.115.

56. Maeda RK, Sitnik JL, Frei Y, Prince E, Gligorov D, Wolfner MF, et al. The lncRNA *male-specific abdominal* plays a critical role in Drosophila accessory gland development and male fertility. *PLoS Genet.* 2018;14: e1007519. doi: 10.1371/journal.pgen.1007519.

57. Li M, Wen S, Guo X, Bai B, Gong Z, Liu X, et al. The novel long non-coding RNA *CRG* regulates Drosophila locomotor behavior. *Nucleic Acids Res*. 2012;40: 11714-27. doi: 10.1093/nar/gks943.

58. Bender W. MicroRNAs in the Drosophila bithorax complex. *Genes Dev*. 2008;22: 14-9. doi: 10.1101/gad.1614208.

59. Dai H, Chen Y, Chen S, Mao Q, Kennedy D, Landback P, et al. The Evolution of Courtship Behaviors through the Origination of a New Gene in Drosophila. *PNAS*. 2008;105: 7478-83. doi: 10.1073/pnas.0800693105.

60. Aprea J, Prenninger S, Dori M, Ghosh T, Monasor LS, Wessendorf E, et al. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J*. 2013;32: 3145-60. doi: 10.1038/emboj.2013.245.

61. Lee H, Zhang Z, Krause HM. Long Noncoding RNAs and Repetitive Elements: Junk or Intimate Evolutionary Partners? *Trends Genet*. 2019;35: 892-902. doi: 10.1016/j.tig.2019.09.006.

62. Amândio AR, Necsulea A, Joye E, Mascrez B, Duboule D. *Hotair* is dispensible for mouse development. *PLoS Genet*. 2016;12: e1006232. doi: 10.1371/journal.pgen.1006232.

63. Bell CC, Amaral PP, Kalsbeek A, Magor GW, Gillinder KR, Tangermann P, et al. The *Evx1/Evx1as* gene locus regulates anterior-posterior patterning during gastrulation. *Sci Rep.* 2016;6: 26657. doi: 10.1038/srep26657.

64. Han X, Luo S, Peng G, Lu JY, Cui G, Liu L, et al. Mouse knockout models reveal largely dispensable but context-dependent functions of lncRNAs during development. *J Mol Cell Biol.* 2018;10: 175-8. doi: 10.1093/jmcb/mjy003.

65. Majumder S, Hadden MJ, Thieme K, Batchu SN, Niveditha D, Chowdhury S, et al. Dysregulated expression but redundant function of the long non-coding RNA HOTAIR in diabetic kidney disease. *Diabetologia*. 2019;62: 2129-42. doi: 10.1007/s00125-019-4967-1.

66. Li C, Shen C, Shang X, Tang L, Xiong W, Ge H, et al. Two novel testis-specific long noncoding RNAs produced by 1700121C10Rik are dispensable for male fertility in mice. *J Reprod Dev.* 2020;66: 57-65. doi: 10.1262/jrd.2019-104.

67. Zhu Y, Lin Y, He Y, Wang H, Chen S, Li Z, et al. Deletion of *lncRNA5512* has no effect on spermatogenesis and reproduction in mice. *Reprod Fertil Dev*. 2020;32: 706-13. doi: 10.1071/RD19246.

68. Kok F, Shin M, Ni C, Gupta A, Grosse A, van Impel A, et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell.* 2015;32: 97-108. doi: 10.1016/j.devcel.2014.11.018.

69. Schor IE, Bussotti G, Maleš M, Forneris M, Viales RR, Enright AJ, et al. Non-coding RNA expression, function, and variation during Drosophila embryogenesis. *Curr Biol.* 2018;28: 3547-61.e9. doi: 10.1016/j.cub.2018.09.026.

70. Goudarzi M, Berg K, Pieper LM, Schier AF. Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. *eLife*. 2019;8. doi: 10.7554/eLife.40815.

71. Selleri L, Bartolomei MS, Bickmore WA, He L, Stubbs L, Reik W, et al. A Hox-embedded long noncoding RNA: Is it all hot air? *PLoS Genet.* 2016;12: e1006485. doi: 10.1371/journal.pgen.1006485.

72. Rappaport Y, Falk R, Achache H, Tzur YB. *Linc-20* and *linc-9* do not have compensatory fertility roles in *C. elegans*. *MicroPubl Biol.* 2022; 2022:10.17912/micropub.biology.000524.

73. Kufel J, Grzechnik P. Small nucleolar RNAs tell a different tale. *Trends Genet.* 2019;35: 104-17. doi: 10.1016/j.tig.2018.11.005.

74. Bratkovič T, Božič J, Rogelj B. Functional diversity of small nucleolar RNAs. *Nucleic Acids Res*. 2020;48: 1627-51. doi: 10.1093/nar/gkz1140.

75. Pritykin Y, Brito T, Schupbach T, Singh M, Pane A. Integrative analysis unveils new functions for the Drosophila Cutoff protein in noncoding RNA biogenesis and gene regulation. *RNA*. 2017;23: 1097-109. doi: 10.1261/rna.058594.116.

76. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet*. 2019;20: 89-108. doi: 10.1038/s41576-018-0073-3.

77. Onishi R, Yamanaka S, Siomi MC. piRNA- and siRNA-mediated transcriptional repression in Drosophila, mice, and yeast: new insights and biodiversity. *EMBO Rep.* 2021;22: e53062-n/a. doi: 10.15252/embr.202153062.

78. Leader DP, Krause SA, Pandit A, Davies SA, Dow JA. FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res.* 2018;46: D809-15. doi: 10.1093/nar/gkx976.

79. Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, et al. FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res*. 2021;49: D899-D907. doi: 10.1093/nar/gkaa1026.

80. Morozova TV, Shankar V, MacPherson RA, Mackay TFC, Anholt RRH. Modulation of the Drosophila Transcriptome by Developmental Exposure to Alcohol. *BMC Genomics*. 2022;23:347. doi: https://doi.org/10.1186/s12864-022-08559-9.

81. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The *Drosophila melanogaster* Genetic Reference Panel. *Nature*. 2012;482: 173-78. doi: 10.1038/nature10811.

82. Everett LJ, Huang W, Zhou S, Carbone MA, Lyman RF, Arya GH, et al. Gene expression networks in the Drosophila Genetic Reference Panel. *Genome Res*. 2020;30: 485-96. doi: 10.1101/gr.257592.119.

83. ElMaghraby MF, Andersen PR, Pühringer F, Hohmann U, Meixner K, Lendl T, et al. A heterochromatin-specific RNA export pathway facilitates piRNA production. *Cell*. 2019;178: 964-79.e20. doi: 10.1016/j.cell.2019.07.007.

84. Kneuss E, Munafò M, Eastwood EL, Deumer U, Preall JB, Hannon GJ, et al. Specialization of the Drosophila nuclear export family protein Nxf3 for piRNA precursor export. *Genes Dev*. 2019;33: 1208. doi: 10.1101/716258.

85. Tycowski KT, Steitz JA. Non-coding snoRNA host genes in Drosophila: expression strategies for modification guide snoRNAs. *Eur J Cell Biol*. 2001;80: 119-25. doi: 10.1078/0171-9335-00150.

86. Dieci G, Preti M, Montanini B. Eukaryotic snoRNAs: A paradigm for gene expression flexibility. *Genomics*. 2009;94: 83-8. doi: 10.1016/j.ygeno.2009.05.002.

87. Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, et al. Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev*. 2014;28: 2498-517. doi: 10.1101/gad.246538.114.

88. Herter EK, Stauch M, Gallant M, Wolf E, Raabe T, Gallant P. snoRNAs are a novel class of biologically relevant Myc targets. *BMC Biol*. 2015;13: 25. doi: 10.1186/s12915-015-0132-6.

89. Kanke M, Jambor H, Reich J, Marches B, Gstir R, Ryu YH, et al. *oskar* RNA plays multiple noncoding roles to support oogenesis and maintain integrity of the germline/soma distinction. *RNA*. 2015;21: 1096-109. doi: 10.1261/rna.048298.114.

90. Kenny A, Morgan MB, Mohr S, Macdonald PM. Knock down analysis reveals critical phases for specific *oskar* noncoding RNA functions during Drosophila oogenesis. *G3 (Bethesda)*. 2021;11. doi: 10.1093/g3journal/jkab340.

91. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics*. 2014;196: 961-71. doi: 10.1534/genetics.113.160713.

92. Kondo S, Ueda R. Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. *Genetics*. 2013;195: 715-21. doi: 10.1534/genetics.113.156737.

93. Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, et al. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res*. 2014;24: 1193-208. doi: 10.1101/gr.171546.113.

94. Carvalho GB, Ja WW, Benzer S. Non-lethal PCR genotyping of single Drosophila. *BioTechniques*. 2009;46: 312-14. doi: 10.2144/000113088.

95. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Δ Δ C(T)) method. *Methods*. 2001;25: 402-8. doi: 10.1006/meth.2001.1262.

96. Mackay TFC, Lyman RF, Jackson MS. Effects of *P* element insertions on quantitative traits in Drosophila melanogaster. *Genetics*. 1992;130: 313-32.

97. Drapeau MD, Long AD. The Copulatron, a multi-chamber apparatus for observing Drosophila courtship behaviors. *Drosophila Information Service*;83: 194-6.

98. Gaertner BE, Ruedi EA, McCoy LJ, Moore JM, Wolfner MF, Mackay TFC. Heritable variation in courtship patterns in *Drosophila melanogaster. G3 (Bethesda).* 2015;5: 531-9. doi: 10.1534/g3.114.014811.

99. Morgan TJ, Mackay TFC. Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity*. 2006;96: 232-42. doi: 10.1038/sj.hdy.6800786.

100. Sass TN, MacPherson RA, Mackay TFC, Anholt RRH. High-Throughput Method for Measuring Alcohol Sedation Time of Individual *Drosophila melanogaster*. *J Vis Exp*. 2020. doi: 10.3791/61108.

101. Cichewicz K, Hirsh J. ShinyR-DAM: a program analyzing Drosophila activity, sleep and circadian rhythms. *Commun Biol.* 2018;1: 25. doi: 10.1038/s42003-018-0031-9.

102. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012;9: 676-82. doi: 10.1038/nmeth.2019.

103. Chen S, Huang T, Zhou Y, Han Y, Xu M, Gu J. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics*. 2017;18: 80. doi: 10.1186/s12859-017-1469-3.

104. Bushnell B. BBMap. SOURCEFORGE. 2014.

105. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. In: Mathé E, Davis S, editors. *Statistical Genomics: Methods and Protocols*. New York, NY: Springer New York; 2016. pp. 283-334.

106. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013;41: e108. doi: 10.1093/nar/gkt214.

107. Smid M, Coebergh van den Braak, Robert R. J., van de Werken, Harmen J. G., van Riet J, van Galen A, de Weerd V, et al. Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics*. 2018;19: 236. doi: 10.1186/s12859-018-2246-7.

108. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4: 2. doi: 10.1186/1471-2105-4-2.

109. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 2013;41: D377-86. doi: 10.1093/nar/gks1118.

110. Jia D, Xu Q, Xie Q, Mio W, Deng W. Automatic stage identification of Drosophila egg chamber based on DAPI images. *Sci Rep*. 2016;6: 18850. doi: 10.1038/srep18850.

# CHAPTER SIX

# GENETIC AND GENOMIC ANALYSES OF *DROSOPHILA MELANOGASTER* MODELS OF CHROMATIN MODIFICATION DISORDERS

MacPherson RA, Shankar V, Anholt RRH, Mackay TFC. Genetic and genomic analyses of *Drosophila melanogaster* models of chromatin modification disorders. *Genetics*. 2023. https://doi.org/10.1093/genetics/iyad061

## Author Contribution Statement

I conducted all Drosophila husbandry, performed all experiments and data analyses, and wrote the original draft of the manuscript. Dr. Vijay Shankar assisted with analysis of the sequencing data.

## Introduction

Switch/Sucrose Non-Fermenting (SWI/SNF)-related intellectual disability disorders (SSRIDDs) and Cornelia de Lange syndrome (CdLS) are syndromic neurodevelopmental Mendelian disorders of chromatin modification. SSRIDDs, including Coffin-Siris syndrome (CSS) and Nicolaides-Baraitser syndrome (NCBRS), stem from variants in genes of the *Brahma-Related Gene-1* Associated Factor (BAF) complex, also known as the mammalian SWI/SNF complex (Hoyer *et al*. 2012; Santen *et al*. 2012; Tsurusaki *et al*. 2012; Van Houdt *et al*. 2012; Tsurusaki *et al*. 2014; Hempel *et al*. 2016; Bramswig *et* *al*. 2017; Bogershausen and Wollnik 2018; Vasileiou *et al*. 2018; Gazdagh *et al*. 2019; Machol *et al*. 2019; Zawerton *et al*. 2019). CdLS is associated with variants in genes that encode components of the cohesin complex (Krantz *et al*. 2004; Deardorff *et al*. 2007; Deardorff *et al*. 2012; Gil-Rodriguez *et al*. 2015; Boyle *et al.* 2017; Huisman *et al*. 2017; Olley *et al*. 2018).

SSRIDD patients exhibit neurodevelopmental delay, intellectual disability, hypotonia, seizures, and sparse hair growth, as well as cardiac, digit, and craniofacial anomalies, where the severity and spectrum of affected phenotypes are dependent upon the specific variant or affected gene product (reviewed in Bogershausen and Wollnik 2018; Schrier Vergano *et al*. 2021; Vasko *et al*. 2021). For example, many SSRIDD patients with variants in *ARID1B* tend to have milder phenotypes including normal growth, milder facial gestalt, and no central nervous system (CNS) abnormalities, whereas most variants in *SMARCB1* are associated with more severe phenotypes, including profoundly delayed developmental milestones, seizures, kidney malformations, and CNS abnormalities (Bogershausen and Wollnik, 2018; Schrier Vergano *et al*. 2021). Furthermore, variants in *ARID1B* are associated with SSRIDD, Autism Spectrum disorder, and non-syndromic intellectual disability (Hoyer *et al*. 2012; De Rubeis *et al*. 2014; Iossifov *et al.* 2014; Vissers *et al*. 2016; van der Sluijs *et al*. 2019). Brain malformations, such as agenesis of the corpus callosum, Dandy-Walker malformation, and cerebellar hypoplasia, have also been observed in 20-30% of all patients with variants in the BAF complex (Vasko *et al.*

2022), but are most commonly observed in patients with variants in *SMARCB1* (Bogershausen and Wollnik 2018).

CdLS patients also display a clinical spectrum including intellectual disability, hirsutism, synophrys, and digit, craniofacial, and CNS anomalies (reviewed in Kline *et al*. 2018; Avagliano *et al*. 2020; Selicorni *et al*. 2021). As in SSRIDDs, some phenotypes are more highly associated with a specific gene, but phenotypic severity can vary widely across variants within the same gene. For example, most patients with variants in *SMC1A* show milder developmental delay and intellectual disability compared to their classical *NIPBL*-CdLS counterparts, but about 40% of *SMC1A* patients exhibit severe epileptic encephalopathy and intellectual disability (Jansen *et al*. 2016; Symonds *et al*. 2017; Selicorni *et al.* 2021).CdLS has also been reclassified as a spectrum of cohesinopathies (Van Allen *et al*. 1993; Kline *et al*. 2018). Patients with pathogenic variants in many genes involved in chromatin accessibility and regulation have overlapping symptoms with CdLS (Parenti et al., 2017; Aoi *et al*. 2019; Cucco *et al*. 2020).

*D. melanogaster* is well-suited for modeling human disorders, as large numbers of flies can be raised economically without ethical or regulatory restrictions. Additionally, SSRIDD- and CdLS-associated genes are highly conserved in flies and a wide variety of genetic tools are available to create fly models of human diseases (Hu *et al*. 2011; Perkins *et al*. 2015; Zirin *et al*. 2020). Previous groups have used *D. melanogaster* to investigate SSRIDDs and CdLS and have observed phenotypes relevant to disease presentation in humans, including changes in sleep, brain function, and brain morphology (Pauli *et al*. 2008; Schuldiner *et al*. 2008; Wu *et al*. 2015; Chubak *et al*. 2019). These studies have provided insight into potential disease pathogenesis and suggested that certain subtypes of SSRIDD and CdLS can be modeled in the fly, but they were not performed in controlled genetic backgrounds.

Here, we present behavioral and transcriptomic data on Drosophila models of SSRIDDs and CdLS in a common genetic background. RNAi-mediated knockdown of Drosophila orthologs of SSRIDD- and CdLS-associated genes show gene- and sex-specific changes in brain structure and sensorimotor integration, as well as increased locomotor activity and decreased night sleep. Transcriptomic analyses show distinct differential gene expression profiles for each focal gene.

## Results

### *Drosophila Models of SSRIDDs and CdLS*

We identified Drosophila orthologs of 12 human genes associated with the SSRIDD chromatin remodeling disorders and CdLS with a DIOPT score > 9 and for which TRiP RNAi lines in a common genetic background and without predicted off-target effects were publicly available. Using these criteria, the Drosophila genes *Bap111*, *brm*, *osa*, and *Snr1* are models of SSRIDD-associated genes *ARID1A*, *ARID1B*, *SMARCA2*, *SMARCA4*, *SMARCB1*, and *SMARCE1*; and *Nipped-B*, *SMC1*, *SMC3*, and *vtd* are models of CdLS-associated genes *NIPBL*, *SMC1A*, *SMC3*, and *RAD21* (Table S2).

We obtained *UAS*-RNAi lines generated in the same genetic background for each of the fly orthologs and crossed these RNAi lines to each of three ubiquitous *GAL4* drivers to assess viability (Figure S1). We selected ubiquitous drivers since the human SSRIDD- and CdLS-associated genes and Drosophila orthologs are ubiquitously expressed, and SSRIDD and CdLS patients carry pathogenic variants in all cells. We initially crossed each *UAS*-RNAi line to three ubiquitous *GAL4* drivers (*Actin-GAL4*, *Ubiquitin-GAL4*, and *Ubi156-GAL4*) and assessed viability and degree of gene knockdown in the F1 progeny (Figure S1). *Ubiquitin-GAL4*-mediated gene knockdown resulted in viable progeny in only three of the eleven *UAS*-RNAi lines, with most progeny dying during the embryonic or larval stage (Figure S1). Based on these data, we selected the weak ubiquitin driver *Ubi156-GAL4* (Garlapow *et al*. 2015) and the *UAS-RNAi* lines for *brm*, *osa*, *Snr1*, *SMC1*, *SMC3*, and *vtd* for further study (Table 1). With the exception of *Ubi156>osa* males which had ~15% gene knockdown, RNAi knockdown of all genes ranged from 40-80% (Table S3). Given that SSRIDDs and CdLS are largely autosomal dominant disorders, knockdown models that retain some degree of gene expression are reflective of the genetic landscape of SSRIDD and CdLS patients.

**Table 6.1. Drosophila genes used in fly models.** The table indicates fly genes used in SSRIDD and CdLS fly models, as well as the respective human orthologs and MIM numbers, associated human disease and respective MIM numbers, and DIOPT scores. Human orthologs are only included in the table if the DIOPT score is greater than 9.

| Fly Gene | Human Ortholog(s) | Human Ortholog MIM number(s) | Associated Human Disease | Phenotype MIM Number(s) | DIOPT score |
|---|---|---|---|---|---|
| *brm* | *SMARCA2, SMARCA4* | 600014, 603254 | SSRIDD (NCBRS, CSS 4) | 601358, 614609 | 13, 12 |
| *osa* | *ARID1A, ARID1B* | 603024, 614556 | SSRIDD (CSS 2, CSS 1) | 614607, 135900 | 12, 12 |
| *SMC1* | *SMC1A* | 300040 | Cornelia de Lange syndrome 2 | 300590 | 12 |
| *SMC3* | *SMC3* | 606062 | Cornelia de Lange syndrome 3 | 610759 | 12 |
| *Snr1* | *SMARCB1* | 601607 | SSRIDD (CSS 3) | 614608 | 15 |
| *vtd* | *RAD21, RAD21L1* | 606462, 619533 | Cornelia de Lange syndrome 4 | 614701 | 11, 10 |

### *Effects on Startle Response*

Given the neurological and musculoskeletal clinical findings in SSRIDD, and CdLS patients (Bogershausen and Wollnik 2018; Kline *et al*. 2018; Avagliano *et al*. 2020; Schrier Vergano *et al*. 2021; Selicorni *et al*. 2021; Vasko *et al*. 2022), we assessed startle-induced sensorimotor integration for RNAi of *brm*, *osa*, *Snr1*, *SMC1*, *SMC3*, and *vtd* relative to their control genotype. Almost all genotypes exhibited a decreased startle response across both sexes ($p < 0.02$ for all by-sex by-genotype comparisons to the control, Figure 1A, Table S4). Males with *osa* or *brm* knockdown did not exhibit changes in startle response ($p > 0.05$), and females with *Snr1* knockdown showed an increased startle response ($p < 0.0001$). In the lines where both sexes were affected, we observed more extreme phenotypes in males (Figure 1A).

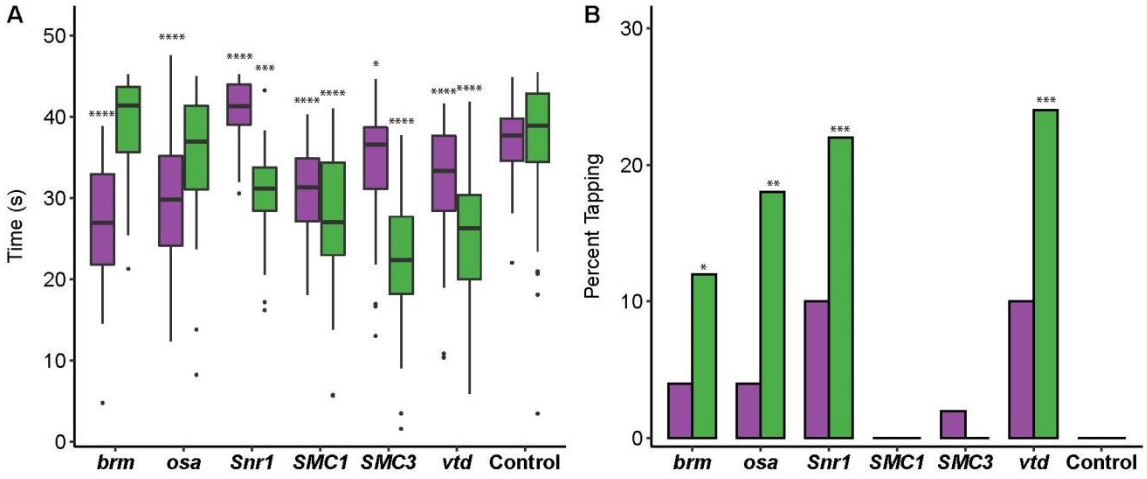


**Figure 6.1. Altered startle response phenotypes in SSRIDD and CdLS fly models.** Startle phenotypes of flies with *Ubi156-GAL4*-mediated RNAi knockdown. (A) Boxplots showing the time, in seconds, spent moving after an initial startle force. Asterisks represent sex-specific pairwise comparisons with the control. (B) Bar graphs showing the percentage of flies that exhibit tapping behavior (see File S1 and S2) following an initial startle stimulus. Females and males are shown in purple and green, respectively. See Table S4 for ANOVAs (A) and Fisher's Exact Tests (B). N = 36-50 flies per sex per line. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

While testing flies for startle response, we noticed that some flies exhibited a specific locomotion phenotype we termed "tapping". Tapping is characterized by repetitive extension and retraction of individual legs as if to walk, but without progressive movement in any direction (File S1). Compared to the control (example shown in File S2), we observed an increase in the number of flies exhibiting tapping behavior in male flies with knockdown of *brm* ($p = 0.0267$), *osa* ($p = 0.0026$), *Snr1* ($p = 0.0005$) and *vtd* ($p = 0.0002$)

(Figure 1B, Table S4). We also observed increases in tapping behavior in females with knockdown of *Snr1* and *vtd* that fall just outside of a significance level of 0.05 ($p = 0.0563$ for both genes); Figure 1B, Table S4). The tapping and startle phenotypes were not evident across all genes associated with a specific disorder.

### *Effects on Sleep and Activity*

We hypothesized that hypotonia and sleep disturbances observed in SSRIDD and CdLS patients (Liu and Krantz 2009; Stavinoha *et al*. 2011; Rajan *et al*. 2012; Zambrelli *et al*. 2016; Bogenshausen and Wollnik 2018; Schrier Vergano *et al*. 2021; Vasko *et al*. 2021) may correspond to changes in activity and sleep in Drosophila models. Sleep disturbances were also observed in a previous Drosophila model of *NIPBL*-CdLS (Wu *et al*. 2015). Therefore, we quantified activity and sleep phenotypes for RNAi-mediated knockdown of *brm*, *osa*, *Snr1*, *SMC1*, *SMC3*, and *vtd*. All RNAi genotypes showed increases in overall spontaneous locomotor activity ($p < 0.02$ for all by-sex by-genotype comparisons to the control, Figure 2A, Table S4). This increase in spontaneous locomotor activity was most pronounced in males with knockdown of *osa* ($p < 0.0001$); this was the only genotype for which males were more active than females (Figure 2A, Table S4). All RNAi genotypes showed decreases in night sleep ($p < 0.0001$ for all by-sex by-genotype comparisons to the control). Flies with knockdown of *osa* (males, $p < 0.0001$; females, $p < 0.0001$) and females with knockdown of *vtd* ($p < 0.0001$) spent about half of the nighttime awake, the least amount of sleep across all flies tested (Figure 2B, Table S4). In addition to increased activity, the Drosophila models of SSRIDDs and CdLS have fragmented sleep: the number of sleep bouts at night was increased for all lines and sexes compared to the control ($p <$ 0.0001 for all by-sex by-genotype comparisons to the control, except *SMC1* males, $p =$ 0.0023, Figure 2C, Table S4).

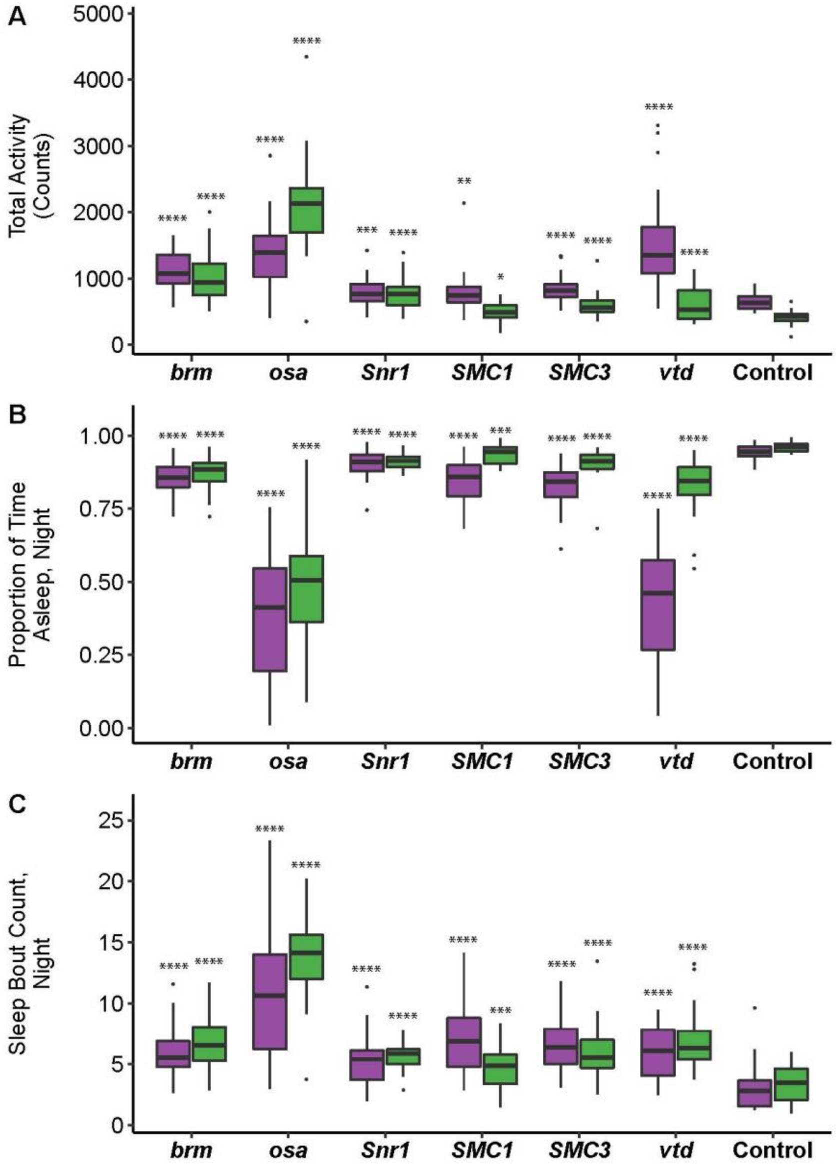


**Figure 6.2. Altered sleep and activity phenotypes in SSRIDD and CdLS fly models.** Boxplots displaying activity and sleep phenotypes of flies with *Ubi156-GAL4*-mediated

RNAi knockdown. (A) total activity; (B) proportion of time spent asleep at night; (C) number of sleep bouts at night. Females and males are shown in purple and green, respectively. N = 18-32 flies per sex per line. See Table S4 for ANOVAs. Asterisks indicate pairwise comparisons of each line to the control, sexes separately. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

### *Effects on Brain Morphology*

To assess changes in brain structure in *brm*, *osa*, *Snr1*, *SMC1*, *SMC3,* and *vtd* RNAi genotypes, we focused on the mushroom body and the ellipsoid body, as prior studies on SSRIDDs in flies have shown changes in mushroom body structure (Chubak *et al*. 2019), and the mushroom body has been linked with regulation of sleep and activity in Drosophila (Joiner *et al*. 2006; Pitman *et al*. 2006; Guo *et al*. 2011; Sitaraman *et al*. 2015). Furthermore, SSRIDD and CdLS patients often present with intellectual disability and CNS abnormalities (Bogershausen and Wollnik 2018; Kline *et al*. 2018; Avagliano *et al*. 2020; Schrier Vergano *et al*. 2021; Selicorni *et al*. 2021; Vasko *et al*. 2022). In the Drosophila brain, the mushroom body mediates experience-dependent modulation of behavior (reviewed in Modi *et al*. 2020), making the mushroom body and the ellipsoid body, which mediates sensory integration with locomotor activity, suitable targets for examining changes in brain structure. We used confocal microscopy to quantify the lengths of both alpha and beta lobes of the mushroom body, as well as the horizontal and vertical lengths of the ellipsoid body (Figures 3A-B). The lengths of these lobes were measured in three dimensions, capturing the natural curvature of the alpha and beta lobes of the mushroom body instead of relying upon a 2D measurement of a 3D object.

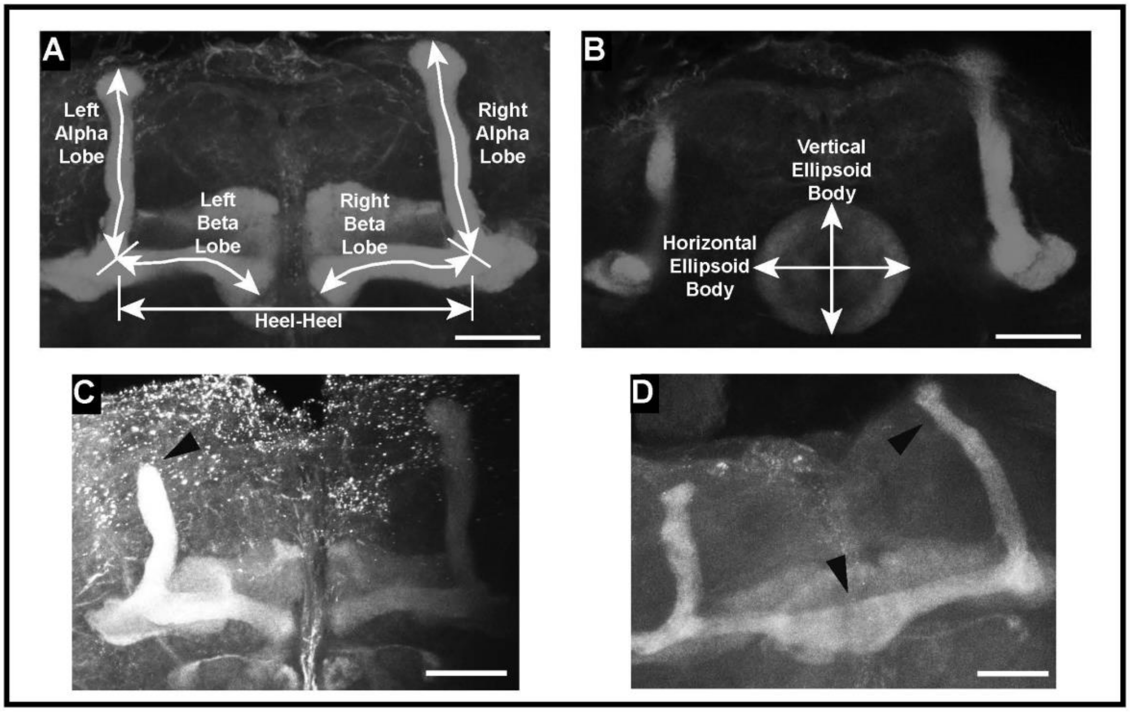

A
Left Alpha Lobe
Right Alpha Lobe
Left Beta Lobe
Right Beta Lobe
Heel-Heel
B
Vertical Ellipsoid Body
Horizontal Ellipsoid Body
C
D

**Figure 6.3. Examples of mushroom body abnormalities in SSRIDD and CdLS fly models.** Images of a wild type mushroom body annotated with measurement descriptors for (A) mushroom body alpha and beta lobes, and heel-heel normalization measurement; and (B) ellipsoid body measurements. Images of select brains from flies with *Ubi156-GAL4*-mediated RNAi knockdown of *osa* showing (C) stunted alpha lobe outgrowth and narrowed alpha lobe head in a female *osa*-deficient fly brain; and (D) beta lobe crossing the midline/fused beta lobes, as well as a skinny alpha lobe in a male *osa*-deficient fly brain. Images shown are z-stack maximum projections from confocal imaging. Triangular arrowheads indicate the abnormalities. The scale bar represents 25 $\mu$M.

We observed sex-specific changes in brain morphology (Figure 3C-D). Females, but not males, showed decreased ellipsoid body dimensions with knockdown of *Snr1* (horizontal, $p = 0.0002$; vertical, $p < 0.0444$, Table S4), while knockdown of *vtd* in females showed decreased alpha ($p = 0.0088$) and beta ($p = 0.0433$) lobe lengths. In addition to sex-specific effects, we observed sexually dimorphic effects; females with knockdown of *brm* showed decreases in alpha lobe and horizontal ellipsoid body length ($p = 0.0409$, $p = 0.0224$, respectively), while *brm* knockdown males showed increases in alpha lobe and horizontal ellipsoid body length ($p = 0.0301$, $p = 0.0305$, respectively; Figure 4, Table S4). Levene's tests for equality of variances indicate that the ellipsoid body measurements have sex-specific unequal environmental variances in some genotypes compared to the control (Figure 4, Table S4). These results show that these models of SSRIDDs and CdLS show morphological changes in the mushroom body and ellipsoid body.

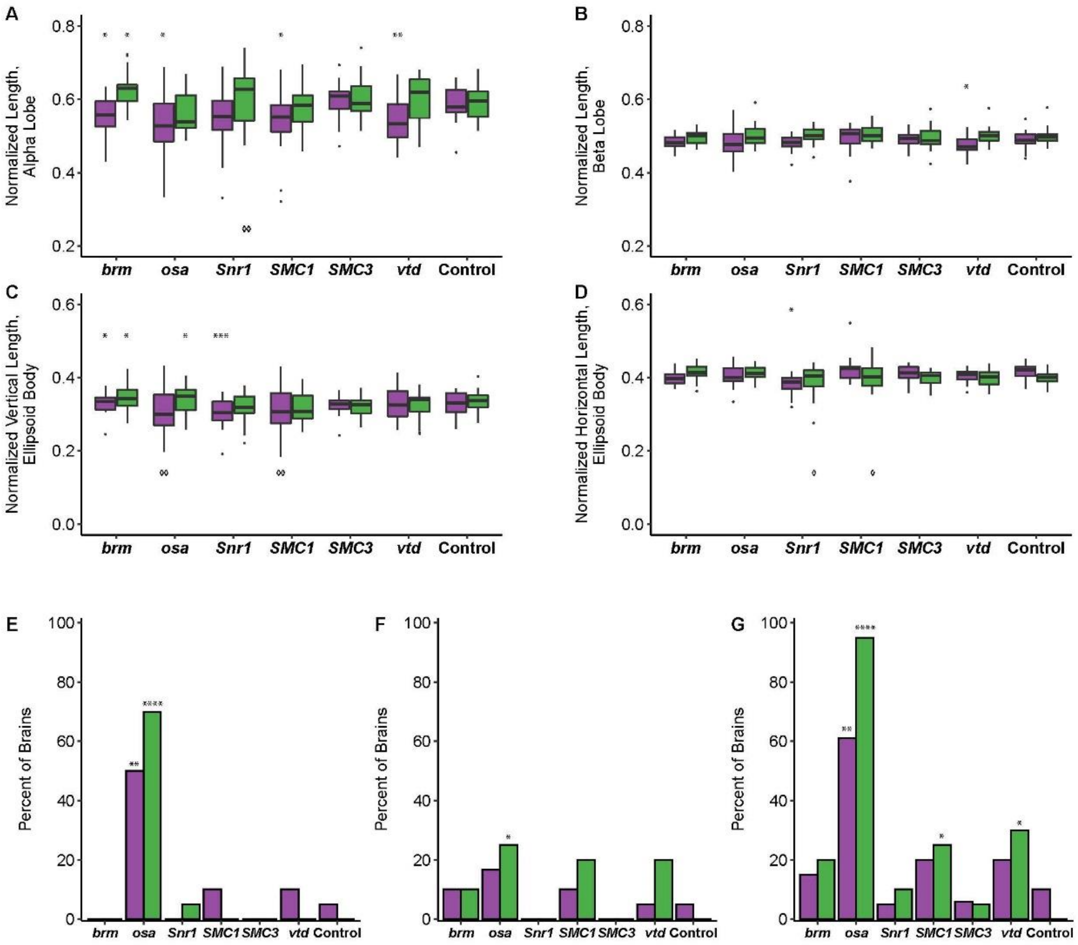


**Figure 6.4. SSRIDD and CdLS fly models show gene-specific changes in mushroom body and ellipsoid body.** Boxplots showing (A) the average alpha lobe and (B) beta lobe length for each brain; (C) ellipsoid body height (vertical direction; dorsal-ventral) and (D) width (left-right; lateral). Bar graphs showing the percentage of brains that (E) have a stunted alpha lobe(s)/narrowed alpha lobe head(s); (F) have a beta lobe(s) crossing the midline, including fused beta lobes; and (G) display one of more of the following defects: skinny alpha lobe, missing alpha lobe, skinny beta lobe, missing beta lobe, stunted alpha lobe/narrowed alpha lobe head, beta lobe crossing the midline/fused beta lobes, extra

projections off of the alpha lobe, extra projections off of the beta lobe. See Figure 3. All brains were dissected from flies with *Ubi156-GAL4*-mediated RNAi knockdown. For panels A-D, brains missing only one alpha or beta lobe are represented by the length of the remaining lobe and brains missing both alpha lobes or both beta lobes were not included in the analyses. For panels E-G, data were analyzed with a Fisher's Exact test, sexes separately. Asterisks (*) and diamonds (panels A-D only; ◊) represent pairwise comparisons of the knockdown line versus the control in ANOVAs or Fisher's Exact tests, and Levene's tests for unequal variances, respectively. See Table S4 for ANOVAs, Fisher's Exact and Levene's Test results. Females and males are shown in purple and green, respectively. N = 17-20 brains per sex per line. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$. ◊: $p < 0.05$, ◊◊: $p < 0.01$.

We also recorded gross morphological abnormalities, such as missing lobes, beta lobes crossing the midline, and impaired/abnormal alpha lobe outgrowth (Figure 3C-D). Although each abnormality was observed across multiple genotypes, only flies with knockdown of *osa* demonstrated consistent brain abnormalities. Male and female *osa* knockdown flies both exhibited an increased number of alpha lobes with impaired outgrowth (males: $p < 0.0001$, females: $p < 0.0025$, Figure 4E, Table S4), and the *osa* knockdown males also showed a significant number of beta lobe midline defects ($p = 0.0471$, Figure 4F, Table S4). Males with knockdown of *SMC1* and *vtd* also showed increased numbers of abnormal brains ($p = 0.0471$, $p = 0.0202$ respectively; Figure 4G, Table S4). Changes in brain morphology are more gene- and sex-dependent than changes in sleep, activity, and startle response.

### *Effects on Genome-wide Gene Expression*

We performed genome-wide analysis of gene expression for the *brm*, *osa*, *Snr1*, *SMC1*, *SMC3*, and *vtd* RNAi genotypes and their control, separately for males and females. We first performed a factorial fixed effects analysis of variance (ANOVA) for each expressed transcript, partitioning variance in gene expression between sexes, lines, and the line by sex interaction for all seven genotypes. We found that 8,481 and 6,490 genes were differentially expressed (FDR < 0.05 for the Line and/or Line×Sex terms, Table S5), for a total of 9,657 unique genes.

*brm*, *osa*, *Snr1* and their human orthologs (Tables 1, S2) are part of the same protein complex (BAF complex in humans, BAP-complex in flies). Therefore, we evaluated whether other BAP complex members *Bap55*, *Bap60*, and *Bap111* (which are orthologous to human BAF complex members *ACTL6A*, *SMARCD1*, and *SMARCE1*, respectively), are differentially expressed in the analysis of all genes. We observed differential expression of strong fly orthologs (DIOPT > 9) of additional BAF complex subunits in the global model and found that *Bap55* and *Bap60* (FDR-corrected Line *p*-values: 0.0123, 0.01306, respectively; Table S5), but not *Bap111*, are differentially expressed. We did not observe differential expression of *Nipped-B* in the global analysis. *Nipped-B* is a member of the fly cohesin complex along with *SMC1*, *SMC3*, and *vtd*, and is orthologous to the human cohesin complex member *NIPBL*.

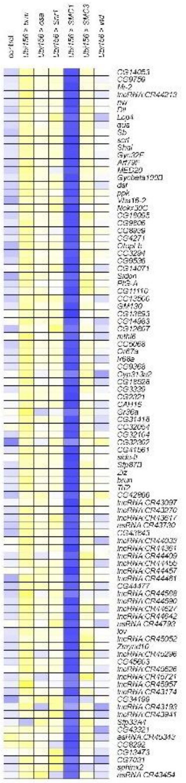
We next performed separate pairwise analyses for SSRIDD-associated fly orthologs and CdLS-associated fly orthologs against the control genotype using the subset of 9,657 unique differentially expressed genes from the full ANOVA model (Tables 2, S5). We also performed these analyses on sexes separately (Tables 2, S5). The number of differentially expressed genes at a given FDR threshold varies across pairwise comparisons and across sexes. For example, females with knockdown of *brm* and *Snr1* have 583 and 3,026 differentially expressed genes (FDR < 0.05), respectively, whereas males with knockdown of these genes have 2,996 and 3,376 differentially expressed genes (FDR < 0.05), respectively (Tables 2, S5). We observed the largest number of differentially expressed genes in flies with knockdown of *Snr1* (Tables 2, S5). At FDR < 0.0005, there were still 1,059 genes differentially expressed in *Snr1* males (Table S5). A greater number of differentially expressed genes are upregulated than downregulated in flies with knockdown of *brm*, *SMC1*, *SMC3*, and *vtd* (Table S5). In contrast, flies with knockdown of *osa* and *Snr1* have a greater number of downregulated genes (Table S5). Flies with knockdown of *Snr1* and *SMC1* had the greatest percentage of differentially expressed genes shared between males and females: 12.2% (698) and 7.6% (348) respectively (Table S6). *Snr1* also had the greatest percent knockdown by RNAi. Only four genes are differentially expressed in all pairwise comparisons of knockdown lines versus the control line, in both males and females; all are computationally predicted genes (Table S6).

**Table 6.2. Differentially expressed gene counts.** The table shows the number of differentially expressed genes (FDR < 0.05) for the Line and/or Line × Sex terms for each pairwise analysis of knockdown vs control, sexes together and sexes separately.

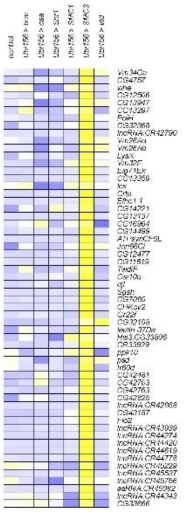
| | **Analysis** | | | |
|---|---|---|---|---|
| | **Both sexes** | | **Females only** | **Males only** |
| **Comparison** | **Line** | **Line×Sex** | **Line** | **Line** |
| *brm* vs. Control | 2808 | 1652 | 583 | 2995 |
| *osa* vs. Control | 2179 | 1059 | 1135 | 1580 |
| *Snr1* vs. Control | 4996 | 3632 | 3026 | 3376 |
| *SMC1* vs. Control | 2714 | 1727 | 2540 | 2395 |
| *SMC3* vs. Control | 1874 | 586 | 2711 | 1161 |
| *vtd* vs. Control | 1998 | 961 | 818 | 1630 |

We performed k-means clustering to examine patterns of co-regulated expression, separately for males (k=8) and females (k=10). We identified the cutoff threshold value for Log2FC by first sorting genes in a descending order of maximal absolute value of Log2FC (Table S7). We fitted lines to roughly linear segments of the generated distribution and designated the cutoff threshold as the Log2FC value of the index at the intersection of the two fitted lines (Figure S2, Table S7). The genes in each cluster are listed in Table S8. Although many clusters reveal gene-specific expression patterns (*e.g.* Cluster F1, F9, F10, Figure 5; Clusters M1, M6, Figure 6), Clusters F7 and F8 show disease-specific patterns, where knockdown of *brm, osa,* and *Snr1* clusters separately from *SMC1*, *SMC3,* and *vtd* (Figure 5). This is not surprising, as *brm, osa,* and *Snr1* are part of the fly BAF complex and models for SSRIDDs, whereas *SMC1*, *SMC3*, and *vtd* are associated with the fly cohesin complex and are models for CdLS. We also observed patterns involving genes from both SSRIDDs and CdLS. Clusters F4 and M3 contain genes upregulated in response to knockdown of *SMC3*, *osa*, and *brm* and downregulated in response to knockdown of *Snr1* and *SMC1* (Figures 5-6) Clusters F5 and M5 contain genes upregulated only in flies with knockdown of *osa* and *Snr1* (Figures 5-6). Notably, many long noncoding RNAs (lncRNAs) feature prominently in many of the male and female clusters (Figures 5-6; Tables S7, S8).

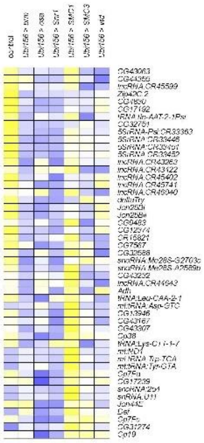
F1



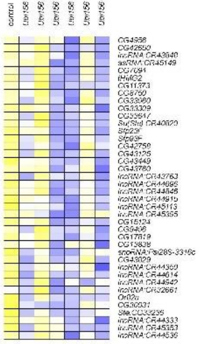
F2



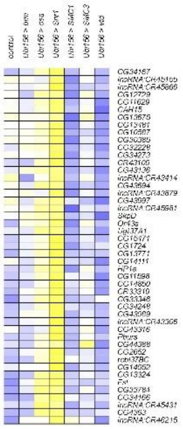
F3



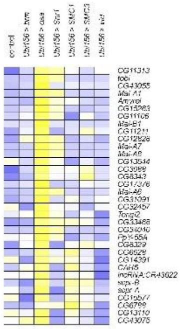
F4



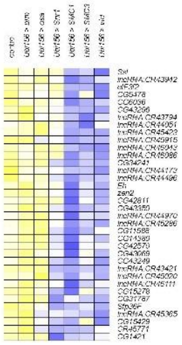
F5



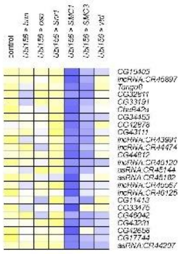
F6



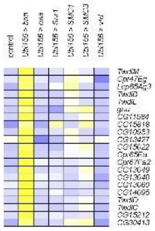
F7



F8



F9



F10

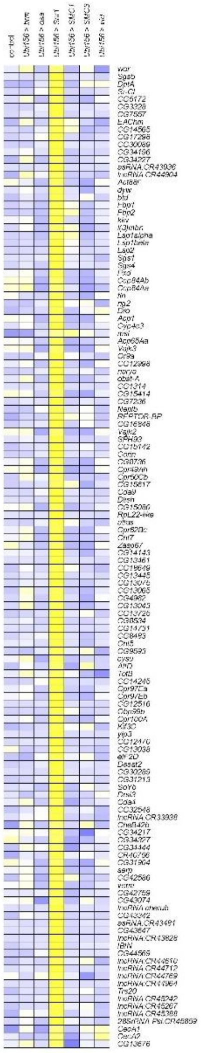


**Figure 6.5. k-means clusters for females.** k-means clusters (k = 10, average linkage algorithm) based on expression patterns of the 535 genes with maximal absolute value of the fold-change in expression, compared to the control. Blue and yellow indicate lower and higher expression, respectively.

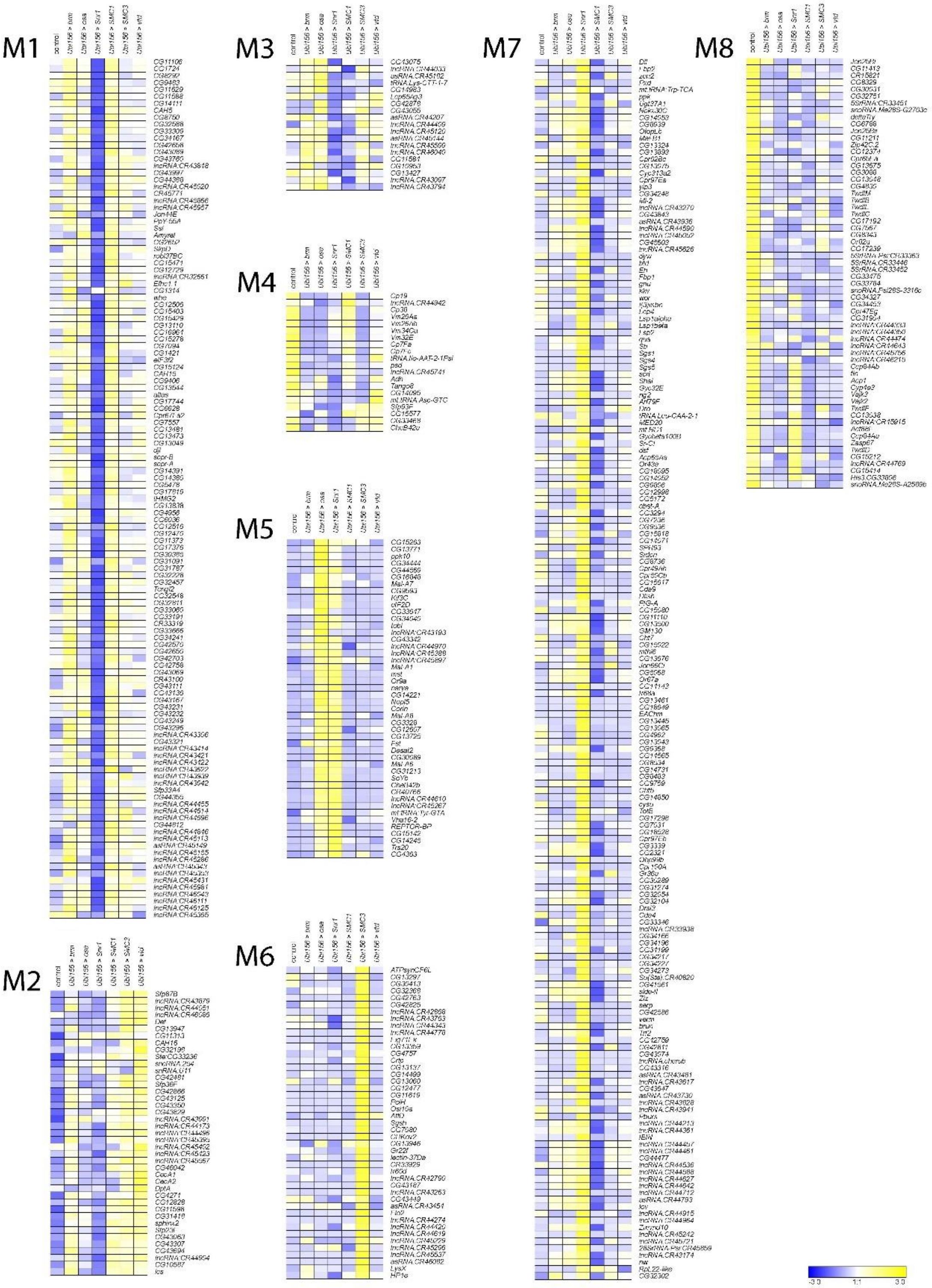

M1
M2
M3
M4
M5
M6
M7
M8
control
Ubi156 > brm
Ubi156 > osa
Ubi156 > Snr1
Ubi156 > SMC1
Ubi156 > SMC3
Ubi156 > vtd
-3.0
1:1
3.0

**Figure 6.6. k-means clusters for males.** k-means clusters (k = 8, average linkage algorithm) based on expression patterns of the 535 genes with maximal absolute value of the fold-change in expression, compared to the control. Blue and yellow indicate lower and higher expression, respectively.

To infer functions of these differentially expressed genes, we performed Gene Ontology (GO) analyses on the top approximately 600 (1000) differentially expressed genes for sexes separately (sexes pooled) (Table S9). These analyses reveal that differentially expressed genes associated with knockdown of CdLS-associated fly orthologs are involved in chromatin organization, regulation and processing of RNA, reproduction and mating behavior, peptidyl amino acid modification, and oxidoreductase activity (Table S9). We also see sex-specific effects, such as muscle cell development in males and neural projection development in females (Table S9). Differentially expressed genes associated with knockdown of SSRIDD-associated fly orthologs in males are involved in mating behavior, cilia development, and muscle contraction, while we see overrepresented ontology terms involved in chromatin modification, mitotic cell cycle, and serine hydrolase activity in females (Table S9). We observed more alignment of GO terms across genes and sexes in the CdLS fly models (*SMC1, SMC3, vtd*) than in SSRIDD fly models (*brm, osa, Snr1*). There were no overrepresented GO terms for females in the CdLS-specific analysis. However, in the 156 genes shared across both sexes and both the SSRIDD and CdLS disease-level analyses, we see an overrepresentation of muscle cell development and actin assembly and organization (Table S9). GO enrichment on k-means clusters does not reveal over-representation of any biological processes, molecular functions or pathways for Clusters F7, F8, F4, F5, and M3 (Table S10). Genes involved in alpha-glucosidase activity are overrepresented in Cluster M5 (Table S10).

We generated Venn diagrams (Figure S3) to display the degree of similarity in differentially expressed genes across analyses, including the 156 genes shared across SSRIDD and CdLS males and females (Table S6). Interestingly, 93% (2689/2907) of genes differentially expressed in a disease-specific analysis of CdLS males were also differentially expressed in CdLS females or in SSRIDD fly models (Table S6). This is in contrast to CdLS females, SSRIDD males, and SSRIDD females, in which about 25% of the differentially expressed genes were specific to a single analysis (Table S6). Approximately 24 and 56 percent of the differentially expressed genes (FDR<0.05) in pairwise comparisons for males and females, respectively, have a predicted human ortholog (DIOPT > 9) (Table S11).

### *Co-Regulated Genes*

We selected a subset of co-regulated genes from gene expression analyses as potential modifiers of the focal genes *brm*, *osa*, and/or *Snr1*. We chose genes that had a significant effect (Line FDR < 0.05) in analyses pooled across sexes, a suggestive effect (Line FDR < 0.1) for each sex separately, a greater than or less than two-fold-change in both sexes, a strong human ortholog (DIOPT > 9), and an available *attp40* TRiP RNAi line (the same genetic background as the focal genes). We increased the FDR threshold to 0.1 for the sex-specific pairwise analyses to account for the decreased power of these analyses compared to those with sexes combined. This resulted in 31 genes (Table S12). We further narrowed our selection by prioritizing genes for further study with potential roles in neurological tissues, metabolism, chromatin, orthologs associated with disease in humans, and computationally predicted genes of unknown function. The six fly genes we selected for further study are *Alp10*, *CG40485*, *CG5877*, *IntS12*, *Mal-A4*, and *Odc1*, which are orthologous to human genes *ALPG*, *DHRS11*, *NRDE2*, *INTS12*, *SLC3A1*, and *ODC1*, respectively (human ortholog with highest DIOPT score listed; Table S12). All six genes tested were co-regulated with *Snr1*, but *CG40485* and *CG5877* were not co-regulated with *osa* and *brm* models of SSRIDDs (Table S6).

For each target gene, we crossed the *UAS-RNAi* line to the *Ubi156-GAL4* driver and performed qRT-PCR to assess the magnitude of reduction in gene expression. All co-regulated genes had reduced expression in both sexes (Table S13). We then assessed the effects of these genes on startle response, sleep, and activity. Knockdown of *Mal-A4, CG5877* and *Alp10* showed changes in startle response times for both sexes (Figure S4A, Table S14). *Mal-A4* demonstrated sexually dimorphic changes in startle response similar to flies with *Snr1* knockdown, as females showed an increase ($p = 0.0215$) and males showed a decrease ($p < 0.0001$) in startle response (Figure S4A, Table S14). We also quantified tapping behavior in these co-regulated genes and found that flies with knockdown of *CG5877* and *Odc1* showed an increase in tapping behavior compared to the control, similar to flies with knockdown of *osa* and *Snr1* (Figure 1B), although we only observed tapping in females with knockdown of *Odc1* (Figure S4B, Table S14; *CG5877* females: $p = 0.0266$, *CG5877* males: $p < 0.0001$; *Odc1* females: $p = 0.0125$).

With the exception of *CG40485*, which showed no changes in sleep or activity for either sex, all male RNAi genotypes had increased nighttime sleep bouts ($p < 0.03$), decreased night sleep ($p < 0.03$), and, with the additional exception of *CG5877* RNAi flies, increased overall activity ($p < 0.006$) (Figure S4, Table S14). Knockdown of *Mal-A4* and *Odc1* also showed increased activity for females ($p = 0.0049$, $p = 0.0044$, respectively). Only knockdown of *CG5877* resulted in increased night sleep for females ($p = 0.0014$) (Figure S4C-D, Table S14). These changes in activity and sleep phenotypes largely parallel those observed for SSRIDD fly models (Figure 2, Table S14).

Based on effects on startle response, tapping behavior, locomotor activity, night sleep, and sleep bouts, none of the phenotypes associated with RNAi of the co-regulated genes exactly matched the phenotypes associated with RNAi of the SSRIDD focal genes in both magnitude and direction. However, three genes (*Mal-A4*, *CG5877*, *Odc1*) exhibited at least one altered phenotype in both sexes (Figure S4). These phenotypic observations provide evidence that *Mal-A4*, *CG5877*, and/or *Odc1* may be interacting with the focal genes of the SSRIDD fly models.

## Discussion

Variants in members of the mammalian SWI/SNF complex (BAF complex) give rise to SSRIDDs, Mendelian disorders with a wide range of phenotypic manifestations, including Coffin-Siris and Nicolaides-Baraitser syndromes (reviewed in Bogershausen and Wollnik 2018; Schrier Vergano *et al.* 2021). The diverse consequences of such variants and variation in penetrance of similar variants in different affected individuals suggest the presence of segregating genetic modifiers. Such modifiers may represent targets for ameliorating therapies or serve as indicators of disease severity, yet they cannot be easily identified in humans due to the limited sample size of individuals with rare disorders. In addition to identifying potential modifiers, Drosophila models can be used to understand underlying molecular effects of variants in chromatin-modification pathways and may aid in discovery of drugs that ameliorate deleterious phenotypic effects.

We used a systematic comparative genomics approach to generate Drosophila models of disorders of chromatin modification, based on the assumption that fundamental elements of chromatin modification are evolutionarily conserved. First, we reduced expression of BAF and cohesin complex orthologs through targeted RNA interference with a *GAL4* driver that induces minimal lethality. We assessed consequences of target gene knockdown on behaviors that mimic those affected in patients with SSRIDDs and CdLS. We used startle behavior, a proxy for sensorimotor integration, and sleep and activity phenotypes to assess the effects of variants in fly orthologues of human genes associated with similar behavioral disorders. These Drosophila models show increased activity, decreased night sleep, and changes in sensorimotor integration. Although we cannot readily recapitulate cognitive developmental defects in Drosophila, these behavioral phenotypes along with brain morphology measurements provide a representative spectrum of behaviors that correlate with human disease phenotypes. We observed gene-specific effects. In addition to showing the largest changes in sleep and activity phenotypes, only *osa* RNAi flies showed stunted mushroom body alpha lobes. Furthermore, only females with knockdown of *Snr1* showed an increase in startle response times. Our neuroanatomical studies focused on morphological changes in the ellipsoid body and mushroom bodies. We cannot exclude effects on other regions in the brain.

Next, we performed whole genome transcriptional profiling to identify co-regulated genes with each focal gene and used stringent filters to identify candidate modifier genes from the larger subset of co-regulated genes. k-means clustering reveals co-regulated genes unique to knockdown of a single protein complex member (Figures S4, S5), yet also shows genes co-regulated in response to knockdown of several, but not all, members of the fly cohesin and SWI/SNF complexes. Gene-specific and cross-disease effects are intriguing, since *brm*, *osa*, and *Snr1* are part of the fly SWI/SNF complex, and *SMC1*, *SMC3*, and *vtd* are part of the fly cohesin complex, yet have widespread gene-specific downstream effects on gene regulation. Upon knockdown of one protein complex member, we did not necessarily find changes in gene expression of other members of the same complex. It is possible that a compensatory mechanism exists that maintains transcript levels of other fly SWI/SNF or cohesin complex members or the focal genes themselves (Dorsett 2009; Raab *et al*. 2017; Van der Vaart *et al*. 2020), such as with *Nipped-B* in a CdLS fly model (Wu *et al*. 2015). Furthermore, the abundance of lncRNAs co-regulated with focal genes (Figures S4, S5, Table S8) is intriguing given the association between lncRNAs, chromatin modification, and changes in gene expression in both flies and humans (Li *et al*. 2019; Statello *et al*. 2021).

*Snr1* is part of the Brahma complex, a core component of the BAP complex and is orthologous to *SMARCB1* (Table S2). *Odc1*, which encodes ornithine decarboxylase, is orthologous to *ODC1* (Table S12), which is associated with Bachmann-Bupp syndrome, a rare neurodevelopmental disorder with alopecia, developmental delay, and brain abnormalities (Prokop *et al*. 2021; Bupp *et al*. 2022). Ornithine decarboxylase is the rate-limiting step of polyamine synthesis, which provides critical substrates for cell proliferation and differentiation (reviewed in Wallace *et al*. 2003; Pegg 2016). Polyamines interact with nucleic acids and transcription factors to modulate gene expression (Watanabe *et al*. 1991; Hobbs and Gilmour 2000; Miller-Fleming *et al*. 2015; Maki *et al*. 2017). *CG5877* is predicted to mediate post-transcriptional gene silencing as part of the spliceosome (Herold *et al*. 2009) and is orthologous to human *NRDE2* (Table S12). *Mal-A4* is predicted to be involved in carbohydrate metabolism (Inomata *et al.* 2019) and is orthologous to *SLC3A1* (Table S12). We observed extensive sexual dimorphism in behavioral phenotypes and transcriptional profiles upon knockdown of SSRIDD- and CdLS-associated genes.

Although we are not aware of transcriptional profiles currently available for SSRIDD patients, RNA sequencing of post-mortem neurons from CdLS patients have shown dysregulation of hundreds of neuronal genes (Weiss *et al*. 2021). RNA sequencing in a *Nipped-B*-mutation fly model of *NIPBL*-CdLS found differential expression of ~2800 genes in the imaginal disc (FDR < 0.05) (Wu *et al*. 2015). Thus, we believe the number of differentially expressed genes upon gene knockdown reported herein is comparable to previous studies.

## Methods

### *Drosophila Genes and Stocks*

We selected SSRIDD-, and CdLS-associated genes with a strong fly ortholog (*Drosophila* RNAi Screening Center Integrative Ortholog Prediction Tool (DIOPT) score > 9) (Hu *et al*. 2011) and a corresponding *attp2* fly line available from the Transgenic RNAi Project (TRiP) (Perkins *et al*. 2015; Zirin *et al*. 2020). We excluded human genes that were orthologous to multiple fly genes to increase the likelihood of aberrant phenotypes upon knockdown of a single fly ortholog. We used *attp40* TRiP lines when assessing phenotypes associated with knockdown of co-regulated genes. We used the $y^{1}$, $sc^{*}$, $v^{1}$, $sev^{21}$; *TRiP2; TRiP3* genotype as the control *UAS* line in all experiments. With the exception of the initial viability screen, we crossed all RNAi lines to a weak ubiquitous *GAL4* driver line, *Ubi156-GAL4* (Garlapow *et al*. 2015). Table S1A lists the Drosophila stocks used.

### *Drosophila Culture*

For all experiments, we maintained flies at a controlled density on standard cornmeal/molasses medium (Genesee Scientific, El Cajon, CA) supplemented with yeast in controlled environmental conditions (25˚C, 50% relative humidity, 12-hour light-dark cycle (lights on at 6 am)). Crosses contained five flies of each sex, with fresh food every 48 hours. After eclosion, we aged flies in mixed-sex vials at a density of 20 flies per vial until used in experiments. We performed experiments on 3-5-day old flies from 8 am to 11 am, unless otherwise noted.

### *Viability*

For the initial viability screen of Drosophila orthologs of SSRIDD- and CdLS-associated genes, we crossed *attp2* TRiP lines and the control line to three ubiquitous *GAL4* driver lines. For the viability screen of co-regulated genes, we crossed *attp40* TRiP lines and the control line to the *Ubi156-Gal4* driver line. From days 0-15, we noted the developmental stage. For stocks that contained balancers, we recorded the associated phenotypic marker in eclosed progeny.

### *Quantitative Real-Time PCR (qRT-PCR)*

For the qRT-PCR analyses of gene expression of RNAi targets of *brm*, *osa*, *SMC1*, *SMC3*, *Snr1*, and *vtd*, we flash froze 3-5-day old whole flies on dry ice and then collected, sexes separately, 30 flies per sample. We stored frozen flies and their extracted RNA at -80˚C. We extracted RNA using the Qiagen RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) by homogenizing tissue with 350μL of RLT Plus Buffer containing β-mercaptoethanol (Qiagen) and DX reagent (Qiagen), using a bead mill at 5m/second for 2 minutes. We quantified RNA with the Qubit RNA BR Assay Kit (ThermoFisher Scientific, Waltham, MA) on a Qubit Fluorometer (ThermoFisher Scientific) according to the manufacturer's specifications. We synthesized cDNA using iScript Reverse Transcription Supermix (Bio-Rad Laboratories, Inc., Hercules, CA) according to the manufacturer's instructions. We quantified expression using quantitative real-time PCR with SYBR$^{TM}$ Green PCR Master Mix (ThermoFisher Scientific), according to manufacturer specifications, but with a total reaction volume of 20μL. We used three biological and three technical replicates per sample and calculated percent knockdown using the ΔΔ*ct* method (Livak and Schmittgen 2001). Table S1B contains primer sequences used. For the qRT-PCR analyses of gene expression for the co-regulated genes *Alp10*, *CG40485*, *CG5877*, *IntS12*, *Mal-A4*, and *Odc1*, we extracted RNA using the Direct-zol RNA MiniPrep Plus Kit (Zymo Research, Irvine, CA) and homogenized tissue with 350μL of Tri-Reagent, using a bead mill at 5m/second for 2 minutes. We used two technical replicates in the qRT-PCR analyses of co-regulated genes.

### *Startle-Induced Locomotor Response*

We assessed startle response using a variation of a previously described assay (Yamamoto *et al*. 2008). In summary, 36-50 flies per sex per line were placed into individual vials to acclimate 24 hours prior to testing. To standardize the mechanical startle stimulus, we placed a vial housing a single 3-5-day old fly in a chute. Removal of a supporting dowel allows the vial to drop from a height of 42 cm, after which it comes to rest horizontally (Huggett *et al*. 2021). We measured the total time the fly spent moving during a period of 45 s immediately following the drop. We also recorded whether the fly demonstrated a tapping phenotype, a series of leg extensions without forward movement. Time spent tapping was not considered movement for startle calculations.

### *Sleep and Activity*

We used the Drosophila Activity Monitoring System (DAM System, TriKinetics, Waltham, MA) to assess sleep and activity phenotypes. At 1-2 days of age, we placed flies into DAM tubes containing 2% agar with 5% sucrose, sealed with a rubber cap (TriKinetics) and a small piece of yarn. We collected data for 7 days on a 12-hour light-dark cycle, with sleep defined as at least 5 minutes of inactivity. We discarded data from flies that did not survive the entire testing period, leaving 18-32 flies per sex per line for analysis. We processed the raw sleep and activity data using ShinyR-DAM (Cichewicz and Hirsh 2018) and used the resulting output data for statistical analysis.

### *Dissection and Staining of Brains*

We dissected brains from cold-anesthetized flies in cold phosphate buffered saline (PBS), before we fixed the brains with 4% paraformaldehyde (v/v in PBS) for 15 minutes, washed with PAXD buffer (1x PBS, 0.24% (v/v) Triton-X 100, 0.24% (m/v) sodium deoxycholate, and 5% (m/v) bovine serum albumin) three times for 10 minutes each, and then washed three times with PBS. We blocked fixed brains with 5% Normal Goat Serum

(ThermoFisher Scientific; in PAXD) for 1 hour with gentle agitation, then stained with 2-5 μg/mL of Mouse anti-Drosophila 1D4 anti-Fasciclin II (1:4) (Developmental Studies Hybridoma Bank; Iowa City, IA) for 16-20 hours at 4˚C. We washed brains three times with PAXD for 10 minutes and stained them with Goat anti-Mouse IgG-AlexaFluor488 (1:100) (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA) for 4 hours. Then, we washed brains with PAXD three times for 10 minutes each prior to mounting with ProLong Gold (ThermoFisher Scientific). We performed all steps at room temperature with gentle agitation during incubations

### *Brain Measurements*

We analyzed 17-20 brains per sex per line using a Leica TCS SPE confocal microscope. We visualized Z-stacks of each brain using Icy v. 2.2.0.0 (de Chaumont *et al*. 2012).

We measured ellipsoid body height and ellipsoid body width by measuring vertical ellipsoid body length from dorsal to ventral, and horizontal ellipsoid body length from left to right (relative to the fly). We also measured lengths of the mushroom body alpha and beta lobes by drawing a single 3D line (3DPolyLine Tool within Icy) through the center of each lobe, adjusting the position of the line while progressing through the z-stack. We measured alpha lobes from the dorsal end of the alpha lobe to the alpha/beta lobe heel (where the alpha and beta lobes overlap) and beta lobes from the median end of the beta lobe to the alpha/beta lobe heel. We normalized the measurements for each brain using the distance between the left and right heels of the mushroom body (heel-heel distance). We used the average alpha and beta lobe lengths for each brain for subsequent analyses. In the case of one missing alpha or beta lobe, we did not calculate an average and instead, used the length of the remaining lobe for analysis. If both alpha or both beta lobes were missing, we removed that brain for analysis of the missing lobes, but retained it for analysis of the other brain regions.

We also recorded gross morphological abnormalities of the mushroom body alpha and beta lobes, including missing lobe, skinny lobe, extra projections, abnormal alpha lobe outgrowth, and beta lobes crossing the midline for each brain. We selected these phenotypes based on prior studies on gross mushroom body morphology (Zwarts *et al*. 2015; Chubak *et al*. 2019).

### *Statistical Analyses*

Unless noted below, we analyzed all behavioral data and brain morphology data in SAS v3.8 (SAS Institute, Cary, NC) using the "PROC GLM" command according to the Type III fixed effects factorial ANOVA model $Y = \mu + L + S + L \times S + \varepsilon$, where $Y$ is the phenotype, $\mu$ is the true mean, $L$ is the effect of line (*e.g.* RNAi line versus the control), $S$ is the effect of sex (males, females), and $\varepsilon$ is residual error. We performed comparisons between an RNAi line and its control. We also performed additional analyses for each sex separately.

We used a Fisher's Exact test (*fisher.test* in R v3.63) to analyze the proportion of flies tapping during startle experiments, the number of brains with a specific morphological abnormality, and the number of brains with any gross morphological abnormality.

We performed Levene's and Brown-Forsythe's Tests for unequal variances on the same data set used for the analysis of lobe lengths. For both tests, we used the *leveneTest* command ((*car* v3.0-11, Fox and Sanford 2019) in R v3.6.3) to run a global analysis comparing all genotypes as well as pairwise comparisons.

### *RNA Sequencing*

We synthesized libraries from 100ng of total RNA using the Universal RNA-seq with Nuquant + UDI kit (Tecan Genomics, Inc., CA) according to manufacturer recommendations. We converted RNA into cDNA using the integrated DNase treatment and used the Covaris ME220 Focused-ultrasonicator (Covaris, Woburn, MA) to generate 350bp fragments. We performed ribosomal RNA depletion and bead selection using Drosophila AnyDeplete probes and RNAClean XP beads (Beckman Coulter, Brea, CA), respectively. We purified libraries after 17 cycles of PCR amplification. We measured library fragment sizes on the Agilent Tapestation using the Agilent High Sensitivity DNA 1000 kit (Agilent Technologies) and quantified library concentration using the Qubit 1X dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific). We pooled libraries at 4nM and loaded them onto an Illumina S1 flow cell (Illumina, Inc., San Diego, CA) for paired-end sequencing on a NovaSeq6000 (Illumina, Inc., San Diego, CA). We sequenced three biological replicates of pools of 30 flies each per sex per line. We sequenced each sample to a depth of ~30 million total reads; we resequenced samples with low read depth (<8 million uniquely mapped reads).

We used the default Illumina BaseSpace NovaSeq sequencing pipeline to demultiplex the barcoded sequencing reads. We then merged S1 flow cell lanes, as well as reads from different runs. We filtered out short and low-quality reads using the *AfterQC* pipeline (v0.9.7) (Chen *et al*. 2017) and quantified remaining levels of rRNA via the bbduk command (Bushnell 2014). We aligned reads to the reference genome (*D. melanogaster* v6.13) using GMAP-GSNAP (Wu *et al*. 2016) and counted these unique alignments to Drosophila genes using the featurecounts pipeline from the Subread package (Liao *et al*. 2013). We excluded genes with a median expression across all samples of less than 3 and genes where greater than 25% of the samples had a counts value of 0. We then normalized the data based on gene length and library size using GeTMM (Smid *et al*. 2018) prior to differential expression analysis.

### *Differential Expression Analyses*

We performed multiple analyses for differential expression in SAS (v3.8; Cary, NC) using the "PROC glm" command. We first performed a fixed effects factorial ANOVA model $Y = \mu + L + S + L \times S + \varepsilon$, where Line ($L$, all RNAi and control genotypes) and Sex ($S$) are cross-classified main effects and Line×Sex ($L \times S$) is the interaction term, $Y$ is gene expression, $\mu$ is the overall mean, and $\varepsilon$ is residual error. We then performed the same analyses only for genes associated with SSRIDDs or for CdLS; i.e., 9,657 genes that were significantly differentially expressed (FDR < 0.05 for the Line and/or Line×Sex terms) in the full model. We ran the ANOVA model for each RNAi genotype compared to the control. Finally, we ran ANOVAs ($Y = \mu + L + \varepsilon$) separately for males and females for the disease-specific and individual RNAi analyses.

### *Gene Ontology and k-means Clustering Analyses*

We performed Gene Ontology (GO) statistical overrepresentation analyses on the top 1,000 differentially expressed genes for the Line term (GO Ontology database released 2022-03-22, Pantherdb v16.0 (Mi *et al*. 2013; Thomas *et al*. 2022)) in each disease-specific and pairwise analysis for GO Biological Process, Molecular Function, and Reactome Pathway terms. For the analyses performed on sexes separately, we used the top 600 differentially expressed genes based on the significance of the Line term. The numbers of differentially expressed genes used in GO enrichment gave maximal GO enrichment with minimal redundancy compared to other numbers of differentially expressed genes.

We performed k-means clustering (average linkage algorithm), sexes separately, on Ge-TMM normalized least squares means of 533 genes that had the highest Log2 fold change (FC) in expression. We identified the cutoff threshold value for Log2FC by first sorting genes in a descending order of maximal absolute value of Log2FC, then fitted lines to roughly linear segments of the generated distribution and designated the cutoff threshold as the Log2FC value of the index at the intersection of the two fitted lines. We used hierarchical clustering (Average Linkage algorithm, WPGMA) to determine the approximate number of natural clusters, then performed clustering with varying values of k to determine the largest number of unique, but not redundant, expression patterns. We also performed GO statistical overrepresentation analyses on genes in each k-means cluster (GO Ontology database released 2022-07-01, Pantherdb v17.0 (Mi *et al*. 2013; Thomas *et al*. 2022)) in each disease-specific and pairwise analysis for GO Biological Process, Molecular Function, and Reactome Pathway terms.

*Data Availability Statement*

All high throughput sequencing data are deposited in GEO GSE213763.

Raw behavioral data, qPCR data, and coding scripts are available on GitHub at https://github.com/rebeccamacpherson/Dmel_models_CSS_NCBRS_CdLS. All *UAS*-RNAi lines used in this study are available at the Bloomington Drosophila Stock Center, except the ubiquitous RNAi driver *Ubi156-GAL4* and the double RNAi lines, which are available upon request.

## Acknowledgements

We thank Dr. Lakshmi Sunkara for assistance with RNA sequencing, Marion R. Campbell III, Miller Barksdale, and Rachel C. Hannah for technical assistance with behavioral assays and brain dissections. We thank Dr. Joshua Walters for helping create Figure 3 and helping dissect brains, and Dr. Richard Steet at the Greenwood Genetic Center for suggestions. We thank Katelynne Collins and Tori Gyorey for assistance with the RNAi studies. We thank the TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing transgenic RNAi fly stocks used in this study.

This work was funded by NIH grants R01 GM128974 and P20 GM139769 to TFCM and RRHA, and F31 HD106719 to RAM.

## References

Aoi H, Mizuguchi T, Ceroni JR, Kim VEH, Furquim I, Honjo RS, Iwaki T, Suzuki T, Sekiguchi F, Uchiyama Y, et al. 2019. Comprehensive genetic analysis of 57 families with clinically suspected Cornelia de Lange syndrome. *J Hum Genet.* 64(10):967-78.

Avagliano L, Parenti I, Grazioli P, Di Fede E, Parodi C, Mariani M, Kaiser FJ, Selicorni A, Gervasini C, Massa V. 2020. Chromatinopathies: A focus on Cornelia de Lange syndrome. *Clin Genet.* 97(1):3-11.

Bögershausen N, Wollnik B. 2018. Mutational landscapes and phenotypic spectrum of SWI/SNF-related intellectual disability disorders. *Front Mol Neurosci.* 11:252.

Boyle MI, Jespersgaard C, Nazaryan L, Bisgaard A-, Tümer Z. 2017. A novel *RAD21* variant associated with intrafamilial phenotypic variation in Cornelia de Lange syndrome – review of the literature. *Clin Genet.* 91(4):647-9.

Bramswig NC, Caluseriu O, Lüdecke H-, Bolduc FV, Noel NCL, Wieland T, Surowy HM, Christen H-, Engels H, Strom TM, et al. 2017. Heterozygosity for *ARID2* loss-of-function mutations in individuals with a Coffin–Siris syndrome-like phenotype. *Hum Genet.* 136(3):297-305.

Bupp C, Michale J, VanSickle E, Rajasekaran S, Bachmann AS. 2022. Bachmann-Bupp syndrome. In: GeneReviews ®. Adam MP, Everman DB, Mirzaa GM et al., editors. Seattle, WA: University of Washington, Seattle.

Burns RB, Dobson CB. 1981. Standard error of the difference between means. In: *Experimental Psychology*. Springer, Dordrecht.

Bushnell B. BBMap. SOURCEFORGE. 2014.

Chen S, Huang T, Zhou Y, Han Y, Xu M, Gu J. 2017. AfterQC: Automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics.* 18(Suppl 3):80.

Chubak MC, Nixon KCJ, Stone MH, Raun N, Rice SL, Sarikahya M, Jones SG, Lyons TA, Jakub TE, Mainland RLM, et al. 2019. Individual components of the SWI/SNF chromatin remodelling complex have distinct roles in memory neurons of the Drosophila mushroom body. *Dis Model Mech.* 12(3):dmm037325.

Cichewicz K, Hirsh J. 2018. ShinyR-DAM: A program analyzing Drosophila activity, sleep and circadian rhythms. *Commun Biol.* 1(1):25.

Cucco F, Sarogni P, Rossato S, Alpa M, Patimo A, Latorre A, Magnani C, Puisac B, Ramos FJ, Pié J, et al. 2020. Pathogenic variants in *EP300* and *ANKRD11* in patients with phenotypes overlapping Cornelia de Lange syndrome. *Am J Med Genet A.* 182(7):1690-6.

de Chaumont F, Dallongeville S, Chenouard N, Hervé N, Pop S, Provoost T, Meas-Yedid V, Pankajakshan P, Lecomte T, Le Montagner Y, et al. 2012. Icy: An open bioimage informatics platform for extended reproducible research. *Nat Methods.* 9(7):690-6.

De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, et al. 2014. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature.* 515(7526):209-15.

Deardorff MA, Bando M, Nakato R, Watrin E, Itoh T, Minamino M, Saitoh K, Komata M, Katou Y, Clark D, et al. 2012. *HDAC8* mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature.* 489:313-7.

Deardorff MA, Kaur M, Yaeger D, Rampuria A, Korolev S, Pie J, Gil-Rodríguez C, Arnedo M, Loeys B, Kline AD, et al. 2007. Mutations in cohesin complex members *SMC3* and *SMC1A* cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. *Am J Hum Genet.* 80(3):485-94.

Dorsett D. 2009. Cohesin, gene expression and development: Lessons from Drosophila. *Chromosome Res.* 17(2):185-200.

Fox J, Weisberg S. 2019. *An R companion to applied regression*. Third ed. Thousand Oaks, CA: Sage.

Garlapow M, Huang W, Yarboro M, Peterson K, Mackay T. 2015. Quantitative genetics of food intake in *Drosophila melanogaster*: e0138129. *PLoS One.* 10(9).

Gazdagh G, Blyth M, Scurr I, Turnpenny PD, Mehta SG, Armstrong R, McEntagart M, Newbury-Ecob R, Tobias ES, Joss S. 2019. Extending the clinical and genetic spectrum of *ARID2* related intellectual disability. A case series of 7 patients. *Eur J Med Genet.* 62(1):27-34.

Gil-Rodríguez MC, Deardorff MA, Ansari M, Tan CA, Parenti I, Baquero-Montoya C, Ousager LB, Puisac B, Hernández-Marcos M, Teresa-Rodrigo ME, et al. 2015. De novo heterozygous mutations in *SMC3* cause a range of Cornelia de Lange syndrome-overlapping phenotypes. *Hum Mutat.* 36(4):454-62.

Guo F, Yi W, Zhou M, Guo A. 2011. Go signaling in mushroom bodies regulates sleep in Drosophila. *Sleep.* 34(3):273-81.

Hempel A, Pagnamenta AT, Blyth M, Mansour S, McConnell V, Kou I, Ikegawa S, Tsurusaki Y, Matsumoto N, Lo-Castro A, et al. 2016. Deletions and de novo mutations of *SOX11* are associated with a neurodevelopmental disorder with features of Coffin–Siris syndrome. *J Med Genet.* 53(3):152-62.

Herold N, Will CL, Wolf E, Kastner B, Urlaub H, Lührmann R. 2009. Conservation of the protein composition and electron microscopy structure of drosophila melanogaster and human spliceosomal complexes. *Mol Cell Biol.* 29(1):281-301.

Hobbs CA, Gilmour SK. 2000. High levels of intracellular polyamines promote histone acetyltransferase activity resulting in chromatin hyperacetylation. *J Cell Biochem.* 77(3):345-60.

Hoyer J, Ekici A, Endele S, Popp B, Zweier C, Wiesener A, Wohlleber E, Dufke A, Rossier E, Petsch C, et al. 2012. Haploinsufficiency of *ARID1B*, a member of the SWI/SNF-A chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am J Hum Genet.* 90(3):565-72.

Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE. 2011. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics.* 12(1):357.

Huggett SB, Hatfield JS, Walters JD, McGeary JE, Welsh JW, Mackay TFC, Anholt RRH, Palmer RHC. 2021. Ibrutinib as a potential therapeutic for cocaine use disorder. *Transl Psychiatry.* 11(1):623.

Huisman S, Mulder PA, Redeker E, Bader I, Bisgaard A, Brooks A, Cereda A, Cinca C, Clark D, Cormier-Daire V, et al. 2017. Phenotypes and genotypes in individuals with *SMC1A* variants. *Am J Med Genet A.* 173A(8):2108-25.

Inomata N, Takahasi KR, Koga N. 2019. Association between duplicated maltase genes and the transcriptional regulation for the carbohydrate changes in *Drosophila melanogaster. Gene.* 686:141-145.

Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, et al. 2014. The contribution of de novo coding mutations to autism spectrum disorder. *Nature.* 515(7526):216-21.

Jansen S, Kleefstra T, Willemsen MH, de Vries P, Pfundt R, Hehir-Kwa JY, Gilissen C, Veltman JA, de Vries BBA, Vissers LELM. 2016. De novo loss-of-function mutations in X-linked *SMC1A* cause severe ID and therapy-resistant epilepsy in females: Expanding the phenotypic spectrum. *Clin Genet.* 90(5):413-9.

Joiner WJ, Crocker A, White BH, Sehgal A. 2006. Sleep in Drosophila is regulated by adult mushroom bodies. *Nature.* 441(7094):757-60.

Kline AD, Moss JF, Selicorni A, Bisgaard A, Deardorff MA, Gillett PM, Ishman SL, Kerr LM, Levin AV, Mulder PA, et al. 2018. Diagnosis and management of Cornelia de Lange syndrome: First international consensus statement. *Nat Rev Genet.* 19(10):649-66.

Krantz ID, McCallum J, DeScipio C, Kaur M, Gillis LA, Yaeger D, Jukofsky L, Wasserman N, Bottani A, Morris CA, et al. 2004. Cornelia de Lange syndrome is caused by mutations in *NIPBL*, the human homolog of *Drosophila melanogaster Nipped-B*. *Nat Genet.* 36(6):631-5.

Li K, Tian Y, Yuan Y, Fan X, Yang M, He Z, Yang D. 2019. Insights into the functions of lncRNAs in Drosophila. *Int J Mol Sci*. 20(18):4646.

Liao Y, Smyth GK, Shi W. 2013. The subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* 41(10):e108.

Liu J, Krantz ID. 2009. Cornelia de Lange syndrome, cohesin, and beyond. *Clin Genet.* 76(4):303-14.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Δ Δ C(T)) method. *Methods.* 25(4):402-8.

Machol K, Rousseau J, Ehresmann S, Garcia T, Nguyen TTM, Spillmann RC, Sullivan JA, Shashi V, Jiang Yh, Stong N, et al. 2019. Expanding the spectrum of BAF-related disorders: De novo variants in *SMARCC2* cause a syndrome with intellectual disability and developmental delay. *Am J Hum Genet.* 104(1):164-78.

Maki K, Shibata T, Kawabata S. 2017. Transglutaminase-catalyzed incorporation of polyamines masks the DNA-binding region of the transcription factor relish. *J Biol Chem.* 292(15):6369-80.

Mi H, Muruganujan A, Thomas PD. 2013. PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 41(Database issue):D377-86.

Miller-Fleming L, Olin-Sandoval V, Campbell K, Ralser M. 2015. Remaining mysteries of molecular biology: The role of polyamines in the cell. *J Mol Biol.* 427(21):3389-406.

Modi MN, Shuai Y, Turner GC. 2020. The Drosophila mushroom body: From architecture to algorithm in a learning circuit. *Annu Rev Neurosci.* 43(1):465-84.

Olley G, Ansari M, Bengani H, Grimes GR, Rhodes J, von Kriegsheim A, Blatnik A, Stewart FJ, Wakeling E, Carroll N, et al. 2018. *BRD4* interacts with *NIPBL* and *BRD4* is mutated in a Cornelia de Lange-like syndrome. *Nat Genet.* 50(3):329-32.

Parenti I, Teresa-Roigo ME, Pozojevic J, Gil SR, Bader I, Braunholz D, Bramswig NC, Gervasini C, Larizza L, Pfeiffer L, et al. 2017. Mutations in chromatin regulators functionally link Cornelia de Lange syndrome and clinically overlapping phenotypes. *Hum Genet.* 136(3):307-20.

Pauli A, Althoff F, Oliveira RA, Heidmann S, Schuldiner O, Lehner CF, Dickson BJ, Nasmyth K. 2008. Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. *Dev Cell.* 14(2):239-51.

Pegg AE. 2016. Functions of polyamines in mammals. *J Biol Chem.* 291(29):14904-12.

Perkins LA, Holderbaum L, Tao R, Hu Y, Sopko R, McCall K, Yang-Zhou D, Flockhart I, Binari R, Shim H, et al. 2015. The transgenic RNAi project at Harvard medical school: Resources and validation. *Genetics.* 201(3):843-52.

Pitman J, McGill J, Keegan K, Allada R. 2006. A dynamic role for the mushroom bodies in promoting sleep in Drosophila. *Nature.* 441(7094):753-6.

Prokop JW, Bupp CP, Frisch A, Bilinovich SM, Campbell DB, Vogt D, Schultz CR, Uhl KL, VanSickle E, Rajasekaran S, et al. 2021. Emerging role of *ODC1* in neurodevelopmental disorders and brain development. *Genes.* 12(4):470.

Raab JR, Runge JS, Spear CC, Magnuson T. 2017. Co-regulation of transcription by BRG1 and BRM, two mutually exclusive SWI/SNF ATPase subunits. *Epigenetics Chromatin.* 10(1):62.

Rajan R, Benke JR, Kline AD, Levy HP, Kimball A, Mettel TL, Boss EF, Ishman SL. 2012. Insomnia in Cornelia de Lange syndrome. *Int J Pediatr Otorhinolaryngol.* 76(7):972-5.

Santen G, Aten E, Sun Y, Almomani R, Gilissen C, Nielsen M, Kant SG, Snoeck IN, Peeters E, Hilhorst-Hofstee Y, et al. 2012. Mutations in SWI/SNF chromatin remodeling complex gene *ARID1B* cause Coffin-Siris syndrome. *Nat Genet.* 44(4):379-80.

Schrier Vergano S, Santen G, Wieczorek D, Wollnik B, Matsumo N, Deardorff MA. 2013. Coffin-Siris syndrome. In: GeneReviews ®. Adam MP, Everman DB, Mirzaa GM, et al., editors. Seattle, WA: University of Washington, Seattle.

Schuldiner O, Berdnik D, Levy JM, Wu JS, Luginbuhl D, Gontang AC, Luo L. 2008. piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Dev Cell.* 14(2):227-38.

Selicorni A, Mariani M, Lettieri A, Massa V. 2021. Cornelia de Lange syndrome: From a disease to a broader spectrum. *Genes.* 12(7):1075.

Sitaraman D, Aso Y, Rubin GM, Nitabach MN. 2015. Control of sleep by dopaminergic inputs to the Drosophila mushroom body. *Front Neural Circuits.* 9:73.

Smid M, Coebergh van den Braak, Robert, van de Werken H, van Riet J, Galen A, Weerd V, Daane M, Bril S, Lalmahomed Z, Kloosterman WP, et al. 2018. Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics.* 19(1):236.

Statello L, Guo C, Chen L, Huarte M. 2021. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol*. 22:96-118.

Stavinoha RC, Kline AD, Levy HP, Kimball A, Mettel TL, Ishman SL. 2010. Characterization of sleep disturbance in Cornelia de Lange syndrome. *Int J Pediatr Otorhinolaryngol.* 75(2):215-8.

Symonds JD, Joss S, Metcalfe KA, Somarathi S, Cruden J, Devlin AM, Donaldson A, DiDonato N, Fitzpatrick D, Kaiser FJ, et al. 2017. Heterozygous truncation mutations of the *SMC1A* gene cause a severe early onset epilepsy with cluster seizures in females: Detailed phenotyping of 10 new cases. *Epilepsia.* 58(4):565-75.

Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L, Mi H. 2022. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* 31(1):8-22.

Tsurusaki Y, Koshimizu E, Ohashi H, Phadke S, Kou I, Shiina M, Suzuki T, Okamoto N, Imamura S, Yamashita M, et al. 2014. De novo *SOX11* mutations cause Coffin–Siris syndrome. *Nat Commun.* 5(1):4011.

Tsurusaki Y, Okamoto Nobuhiko, Fukushima Y, Homma T, Kato M, Hiraki Y, Yamagata T, Yano S, Mizuno S, Sakazume S, et al. 2012. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet.* 44(4):376-8.

van Allen MI, Filippi G, Siegel-Bartelt J, Yong S, McGillivray B, Zuker RM, Smith CR, Magee JF, Ritchie S, Toi A, et al. 1993. Clinical variability within Brachmann-de Lange syndrome: A proposed classification system. *Am J Med Genet.* 47(7):947-58.

van der Sluijs, Pleuntje J., Jansen S, Vergano SA, Adachi-Fukuda M, Alanay Y, AlKindy A, Baban A, Bayat A, Beck-Wödl S, Berry K, et al. 2019. The *ARID1B* spectrum in 143 patients: From nonsyndromic intellectual disability to Coffin–Siris syndrome. *Genet Med.* 21(6):1295-307.

van der Vaart A, Godfrey M, Portegijs V, Heuvel S. 2020. Dose-dependent functions of SWI/SNF BAF in permitting and inhibiting cell proliferation in vivo. *Sci Adv.* 6(21):eaay3823.

van Houdt, Jeroen K. J, Nowakowska BA, Sousa SB, van Schaik, Barbera D. C, Seuntjens E, Avonce N, Sifrim A, Abdul-Rahman OA, van den Boogaard, Marie-José H, Bottani A, et al. 2012. Heterozygous missense mutations in *SMARCA2* cause Nicolaides-Baraitser syndrome. *Nat Genet.* 44(4):445-9.

Vasileiou G, Vergarajauregui S, Endele S, Popp B, Büttner C, Ekici AB, Gerard M, Bramswig NC, Albrecht B, Clayton-Smith J, et al. 2018. Mutations in the BAF-complex subunit *DPF2* are associated with Coffin-Siris syndrome. *Am J Hum Genet.* 102(3):468-79.

Vasko A, Schrier Vergano SA. 2022. Language impairments in individuals with Coffin-Siris syndrome. *Front Neurosci.* 15:802583.

Vasko A, Drivas TG, Schrier Vergano SA. 2021. Genotype-phenotype correlations in 208 individuals with Coffin-Siris syndrome. *Genes.* 12(6):937.

Vissers, Lisenka E. L. M., Gilissen C, Veltman JA. 2016. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet.* 17(1):9-18.

Wallace HM, Fraser AV, Hughes A. 2003. A perspective of polyamine metabolism. *Biochem J.* 376(1):1-14.

Watanabe S, Kusama-Eguchi K, Kobayashi H, Igarashi K. 1991. Estimation of polyamine binding to macromolecules and ATP in bovine lymphocytes and rat liver. *J Biol Chem.* 266(31):20803-9.

Weiss FD, Calderon L, Wang Y, Georgieva R, Guo Y, Cvetesic N, Kaur M, Dharmalingam G, Krantz ID, Lenhard B, et al. 2021. Neuronal genes deregulated in Cornelia de Lange syndrome respond to removal and re-expression of cohesin. *Nat Commun.* 12(1):2919.

Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. 2016. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. In: Statistical genomics: Methods and protocols. Mathé E, Davis S, editors. New York, NY: Springer New York. 283 p.

Wu Y, Gause M, Xu D, Misulovin Z, Schaaf CA, Mosarla RC, Mannino E, Shannon M, Jones E, Shi M, et al. 2015. Drosophila *Nipped-B* mutants model Cornelia de Lange syndrome in growth and behavior. *PLoS Genet.* 11(11):e1005655.

Yamamoto A, Zwarts L, Callaerts P, Norga K, Mackay TFC, Anholt RRH. 2008. Neurogenetic networks for startle-induced locomotion in *Drosophila melanogaster*. *Proc Natl Acad Sci USA.* 105(34):12393-8.

Zambrelli E, Fossati C, Turner K, Taiana M, Vignoli A, Gervasini C, Russo S, Furia F, Masciadri M, Ajmone P, et al. 2016. Sleep disorders in Cornelia de Lange syndrome. *Am J Med Genet C.* 172(2):214-21.

Zawerton A, Yao B, Yeager JP, Pippucci T, Haseeb A, Smith JD, Wischmann L, Kühl SJ, Dean JCS, Pilz DT, et al. 2019. De novo *SOX4* variants cause a neurodevelopmental disease associated with mild dysmorphism. *Am J Hum Genet.* 104(2):246-59.

Zirin J, Hu Y, Liu L, Yang-Zhou D, Colbeth R, Yan D, Ewen-Campen B, Tao R, Vogt E, VanNest S, et al. 2020. Large-scale transgenic Drosophila resource collections for loss- and gain-of-function studies. *Genetics.* 214(4):755-67.

Zwarts L, Vanden Broeck L, Cappuyns E, Ayroles JF, Magwire MM, Vulsteke V, Clements J, Mackay TFC, Callaerts P. 2015. The genetic basis of natural variation in mushroom body size in *Drosophila melanogaster*. *Nat Commun.* 6(1):10115.

# CHAPTER SEVEN

# IDENTIFICATION OF CANDIDATE GENETIC MODIFIERS IN A *DROSOPHILA MELANOGASTER* MODEL OF *ARID1B*-RELATED SWI/SNF-RELATED INTELLECTUAL DISABILITY DISORDER

MacPherson, RA, Collins, KM, Macon, SC, Anholt RRH, Mackay TFC.

## Author Contribution Statement

I performed all experiments and data analysis and wrote the original draft of the manuscript. Katelynne Collins and Sarah Macon helped with data collection and animal husbandry and Katelynne Collins also performed the wing imaging.

## Introduction

*ARID1B* (*AT-rich Interaction Domain 1B*) encodes a core structural component of the highly conserved mammalian SWItch/Sucrose Non-Fermentable (SWI/SNF), or *Brahma Related Gene 1* Associated Factor (BAF) complex (He *et al*. 2020). The BAF complex is an adenosine triphosphate-dependent chromatin remodeler with varied composition across development and tissue type. Although the BAF complex has a number of roles throughout the cell, the BAF complex is antagonistic to the chromatin silencing Polycomb repressive complex and is therefore primarily associated with activation of gene expression (Kadoch *et al.* 2017; Centore *et al.* 2020; He *et al.* 2020). BAF complexes are important for long-range chromatin organization, maintenance of pluripotency, DNA repair, cellular differentiation and development of glia, neurons, muscle, and cardiac tissues (Sokpor *et al.* 2017; Michel *et al*. 2018; Cenik and Shilatifard 2021).

Variants in members of the BAF complex are associated with a range of diseases, including cancer, schizophrenia, autism spectrum disorder (ASD), Kleefstra syndrome, SWI/SNF-related intellectual disability disorders (SSRIDDs), and non-syndromic intellectual disability (Sokpor *et al*. 2017; Cenik and Shilatifard 2021). SSRIDDs are a collection of autosomal dominant neurodevelopmental disorders associated with variants in the mammalian SWI/SNF complex (Bögershausen and Wollnik 2018). Although the phenotypic presentation is dependent upon the specific genetic perturbation, SSRIDD patients present with varying degrees of intellectual disability, hypotonia, facial and digit abnormalities, developmental delay, brain abnormalities, and seizures (Schrier Vergano *et al*. 2021; Vasko *et al.* 2021). ASD is a neurodevelopmental disorder characterized by deficits in social communication and interaction, repetitive behaviors and/or sensory sensitivity not otherwise explained by intellectual disability or developmental delay (American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th edition). *ARID* genes, including *ARID1B,* are more tolerant of variation compared to other members of the BAF complex (Kadoch and Crabtree 2015). *ARID1A* is the most commonly mutated BAF subunit in cancers, while *ARID1B* is one of the genes most commonly associated with general intellectual disability, ASD, and SSRIDDs

(Hoyer *et al.* 2012; Grozeva *et al*. 2015; Kadoch and Crabtree 2015; Yuen *et al.* 2017; Bögershausen and Wollnik 2018; Satterstrom *et al*. 2020).

Variants in *ARID1B* are associated with a wide range of clinical presentations, termed *ARID1B*-related disorder (Vergano *et al.* 2019). Some individuals with *ARID1B* variants may present with mild intellectual disability, ASD or even normal cognition, while others may present with a more severe SSRIDD phenotype (Vergano *et al.* 2019; van der Sluijs *et al*. 2019). *ARID1B*-related SSRIDD patients typically have loss-of-function variants and present with hypotonia, intellectual disability, hearing loss, developmental delay, cryptorchidism, coarse facies, nail abnormalities, and behavioral issues (van der Sluijs *et al.* 2019; Vergano *et al.* 2019; Schrier Vergano *et al.* 2021). However, the genotype-phenotype correlation for *ARID1B*-related disorder is poorly understood. Furthermore, due to the diverse roles of the BAF complex and the variety of associated disease presentations, the exact mechanisms of BAF-related disease pathogenesis remain unresolved.

We hypothesize that naturally occurring genetic modifiers may be contributing to the phenotypic spectrum associated with variants in *ARID1B*. However, due to the small number of individuals with pathogenic *ARID1B* variants, there is insufficient statistical power to identify candidate genetic modifiers using human patient data alone. Although mammalian models have identified changes in brain structure and neural development associated with haploinsufficiency of *ARID1B* (Jung *et al*. 2017; Shibutani *et al.* 2017;

Ellegood *et al*. 2021), government regulations and costs become prohibitive when rearing enough individuals to gain sufficient statistical power for identification of genetic modifiers. The *Drosophila melanogaster* model system is well suited to model SSRIDDs; flies are easy to rear in large numbers at low cost and the BAF complex (Brahma-associated protein complex in flies) is also conserved. Candidate genetic modifiers can be identified in Drosophila, using the *Drosophila melanogaster* genetic reference panel (DGRP), a collection of over 1000 fully sequenced, wild-derived inbred lines (Mackay *et al.* 2012; Huang *et al*. 2014; T. Mackay, personal communication, March 2023). A focal variant can be crossed to unique genetic backgrounds from the DGRP, and phenotypic information of interest from resulting progeny can be used in a genome wide association (GWA) analysis, thus mapping naturally occurring variants modifying the effect of the focal variant that are potential candidate genetic modifiers. This approach has been used to identify candidate genetic modifiers for a range of human diseases (Cukier *et al.* 2008; Lavoy *et al.* 2018; Talsness *et al.* 2020) and Drosophila traits (He *et al.* 2016; Palu *et al.* 2019; Özsoy *et al.* 2021).

Previously, we established *D. melanogaster* models for multiple subtypes of SSRIDDs, including *ARID1B*-related SSRIDD (MacPherson *et al*. 2023). Ubiquitous knockdown of *osa*, the fly ortholog of *ARID1B,* is associated with changes in sensorimotor integration and decreased sleep, as well as sex-specific changes in mushroom body structure and expression of genes related to cell cycle, chromatin modification, and muscle contraction (MacPherson *et al.* 2023). Knockdown of *osa* in post-mitotic neurons also results in changes in long-term memory in males and age-dependent changes in mushroom body axons (Chubak *et al*. 2019). In addition to general roles for *osa* in Drosophila cellular differentiation (Hu *et al.* 2021), *osa* is also required for wing disc regeneration (Tian and Smith-Bolton 2021).

Here, we use the Drosophila model system to identify candidate genetic modifiers for *ARID1B*-related SSRIDD. We first perform a small-scale screen using sleep and activity data from flies with ubiquitous knock down of *osa* and show that genetic modifiers for *osa* exist in the DGRP. We then proceed to a large-scale screen, using wing phenotype data from progeny of 392 DGRP lines crossed to flies with wing-specific knockdown of *osa*. These data show that there are naturally occurring genetic modifiers of *osa* in the DGRP. In the future, these data will be used to perform a genome-wide association analysis to identify modifier alleles, genes, and pathways.

## Methods

### *Drosophila Stocks*

*osa* RNAi stocks from the Transgenic RNA interference project (TRiP) (Perkins *et al*. 2015; Zirin *et al*. 2020) were obtained from the Bloomington Drosophila Stock Center (BDSC). We used the *attp40 osa* RNAi line for experiments involving the third chromosome weak ubiquitous *Ubi156 GAL4* driver (Garlapow et al., 2015) and the *attp2 osa* RNA line for experiments involving the *X* chromosome wing-specific *MS1096 GAL4* driver. We used the *$y^1$, $sc^*$, $v^1$, $sev^{21}$; TRiP2; TRiP3* genotype as the control *UAS* line for initial knockdown experiments. The *Ubi156* driver, DGRP 2.0, and DGRP 3.0 lines used were developed in our laboratory (Garlapow *et al*. 2015; Mackay *et al*. 2012, Huang *et al*. 2014; T. Mackay, personal communication, March 2023).

To generate the *Ubi156>osa* and *MS1096>osa* stocks, we used the crossing schemes outlined in Figures S7.1 and S7.2, respectively. Balancer chromosome stocks were isogenized on a CantonS-B background prior to performing these crosses. A list of Drosophila lines, full genotypes, and abbreviations used in this study is in Table S7.1.

### *Drosophila Culture*

We maintained flies at a controlled density in controlled environmental conditions (25˚C, 50% relative humidity, 12-hour light-dark cycle (lights on at 6 am)). All crosses contained five males and five females, with fresh food provided every 48 hours. We aged flies at a density of 25 flies per vial, sexes separately, until used in experiments. Due to supply chain constraints, we used Nutri-Fly™ Molasses Formulation (Genesee Scientific, El Cajon, CA) for all experiments except the DGRP wing screen, for which we used standard cornmeal/molasses medium (Bloomington Drosophila Stock Center Cornmeal, Molasses, Yeast Recipe).

### *Quantitative Real-Time PCR of osa Knockdown*

Unless otherwise specified, all protocols were followed in accordance with manufacturer’s recommendations. We extracted RNA from *Ubi156>*(*attp40*)*osa* samples containing 30 whole flies, sexes separately, flash frozen on dry ice. Flies and extracted RNA were stored at -80 ˚C. RNA was extracted using the Direct-zol RNA MiniPrep Plus Kit (Zymo Research, Irvine, CA); we homogenized tissue with 350µL of Tri Reagent at 5m/second for two minutes using a bead mill. We eluted the RNA in 60µL of water. We quantified the RNA using the Qubit RNA BR Assay kit (ThermoFisher Scientific, Waltham, MA) on a Qubit fluorometer (ThermoFisher Scientific). We generated cDNA using the iScript Reverse Transcription Supermix (Bio-Rad Laboratories, Inc., Hercules, CA).

We used the SYBR™ Green PCR Master Mix (ThermoFisher Scientific) to perform quantitative real-time PCR and quantify gene expression. The total reaction volume was changed to 20µL, with 10µL SYBR™ Green PCR Master Mix, 1µL each of forward (5' - GGAGCATATATCGCAGGACAAT- 3') and reverse (5' - GTGCTCTGCTGACTGACTTCGT - 3') primer, and 25 ng cDNA. We calculated percent knockdown using the ΔΔ*ct* method (Livak and Schmittgen 2001), with three technical replicates for each of three biological replicates.

### *Sleep and Activity in Ubi156>osa Flies*

We crossed *Ubi156-GAL4* males to *attp40 UAS-osa* virgin females. We placed one day old flies into DAM tubes (Trikinetics) and collected data until flies were eight days old. We scored sleep and activity phenotypes for male (*TRiP X; CSB 2/UAS-osa; Ubi156-GAL4/TRiP3*) and female (*CSB X/TRiP X; CSB 2/UAS-osa; Ubi156-GAL4/TRiP3*) flies.

For the control, we crossed *Ubi156-GAL4* males to virgin females from the TRiP control line and scored resulting male (*TRiP X; CSB 2/TRiP 2; Ubi156-GAL4/TRiP3*) and female (*CSB X/TRiP X; CSB 2/TRiP 2; Ubi156-GAL4/TRiP3)* progeny.

### *Sleep and Activity DGRP Screen*

We crossed *CSB X; UAS-osa/CyO; Ubi-156-GAL4/TM3, Sb* males to virgin females from 10 DGRP lines, selected from a subset of maximally homozygous, minimally related, *Wolbachia*-free DGRP lines (Zhou *et al.* 2020). We placed two-to-three-day old flies into DAM tubes (Trikinetics) and collected data until flies were nine to ten days old. We scored sleep and activity phenotypes for male (*DGRP X; DGRP 2/UAS-osa; DGRP 3/Ubi-156-GAL4*) and female (*DGRP X/CSB X; DGRP 2/UAS-osa; DGRP 3/Ubi-156-GAL4*) flies from crosses to each DGRP line. For the control, we crossed *CSB X; TRiP2; Ubi-156-GAL4* males to DGRP females and collected *DGRP X; DGRP 2/TRiP 2; DGRP 3/Ubi-156-GAL4* males and *DGRP X/CSB X; DGRP 2/TRiP 2; DGRP 3/Ubi-156-GAL4* females for phenotyping sleep and activity.

### *Sleep and Activity*

We assessed sleep and activity phenotypes using the Drosophila Activity Monitor (DAM) system (Trikinetics, Waltham, MA). The DAM system quantifies activity by recording the number of times a fly crosses an infrared beam for every minute during the testing period, where sleep is defined as at least 5 minutes of inactivity. We placed one day old flies into DAM tubes (Trikinetics) and collected data until flies were eight days old. We collected data on two replicates of 16 flies per sex per genotype. DAM tubes contained a 5% sucrose, 2% agar solution and a rubber cap (Trikinetics) at one end and capped at the other end with yarn. Flies that died during the testing period were excluded from analysis. Raw data from the DAM system were initially analyzed using ShinyR-DAM (Cichewicz *et al.* 2018). Output data from ShinyR-DAM were downloaded and filtered (*e.g.* day/night, sleep/activity) for statistical analyses.

### *Statistical Analysis of Sleep and Activity Phenotypes*

DAM data were analyzed in SAS (SAS Institute, Cary, NC) v 3.8 in a Type III Analysis of variance (ANOVA).

For the initial *Ubi156>osa* flies, data were analyzed using the “PROC MIXED” command according to the fixed effects ANOVA $Y = \mu + G + S + G\times S + \mathrm{Rep}(G\times S) + \varepsilon$, where $Y$ is the phenotype, $\mu$ is the true mean, $G$ is the fixed effect of genotype (knockdown, control), $S$ is the fixed effect of sex (males, females), $Rep$ is the random effect of replicate, and $\varepsilon$ is residual error. Data were also analyzed for males and females separately.

For the *Ubi156>osa* DGRP screen, data were analyzed using the “PROC MIXED” command according to the mixed-effects ANOVA $Y = \mu + L + G + S + L\times G + L\times S + G\times S + L\times \mathrm{G}\times S + \varepsilon$, where $Y$ is the phenotype, $\mu$ is the true mean, $L$ is the random effect of DGRP line, $G$ is the fixed effect of genotype (knockdown, control), $S$ is the fixed effect of sex (males, females), and $\varepsilon$ is residual error. Data were also analyzed for males and females separately.

### *Crosses for the Wing-Specific DGRP Screen*

We crossed virgin *MS1096-GAL4/FM6, B; TRiP 2; UAS-osa/TM3, Sb* females to males from randomly selected DGRP 2.0 (Mackay *et al*. 2012; Huang *et al*. 2014) and DGRP 3.0 (T. Mackay, personal communication, March 2023) lines. We collected male *MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa*) and female (*DGRP X/MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa*) progeny from these crosses once daily and aged them in vials with fresh food to avoid damage to the wings, sexes and genotypes separately. For controls, we collected *FM6, B; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa* males and *DGRPX/FM6, B; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa* females. At 3-5 days of age we froze 20 flies per sex per genotype per line and stored them at -20 ˚C.

### *Wing Vein Scoring*

For each sex and genotype, we scored the right wings from 15 randomly selected frozen flies for presence of intact wing veins. We counted whether the anterior cross vein (ACV), posterior cross vein (PCV), and longitudinal veins L2, L3, L4, and L5 were intact and assigned each wing an integer ranging from zero to six based on the number of intact veins. Crumpled wings were given a score of zero and wings with all six veins intact were given a score of six. Veins containing gaps or veins that failed to reach the standard endpoints were not considered intact.

For crosses that yielded fewer than 15 progeny per sex per genotype, all emerged flies of the relevant genotype and sex were frozen and right wings scored.

### *Wing Imaging*

Right wings were dissected from where the wing connects with the thorax with a pair of Dumont forceps (VWR International, Radnor, PA). Dissected wings were placed on a glass slide (ThermoFisher) and covered with a cover slip (VWR International) (no mountant was used). Clear nail polish (Sally Hansen, New York, NY) was used to seal each side of the cover slip. Slides were imaged on a Leica M165FC (Leica Microsystems, Wetzlar, Germany) with a DMC6200 camera (Leica Microsystems).

### *Statistical Analysis of Wing Vein Data*

Pearson's correlation coefficient between the sexes was calculated in R (v4.2.3) using the command *cor* from the base R package. We used the "PROC MIXED" command in SAS v3.8 (SAS Institute) to perform an ANOVA on the wing vein data according to the mixed effects model $Y = \mu + L + S + L\times S + \varepsilon$, where $Y$ is the wing vein score, $L$ is the random effect of DGRP line, $S$ is the fixed effect of sex, and $\varepsilon$ is the residual error. We estimated variance components in SAS using the "PROC VARCOMP" command (Restricted maximum likelihood approach) according to the model $Y = \mu + L + S + L\times S + \varepsilon$. We also performed the wing vein ANOVA and estimated variance components for males and females, separately.

We calculated broad sense heritabilities ($H^2$) by adding the genetic variance components for random effect terms and dividing by the sum of all variance components in the model. We calculated broad sense heritabilities of line means as described above, but instead used a value for residual (environmental) variance that was divided by the average sample size.

### *Data Availability Statement*

Raw wing data, qPCR data, and code used for analysis are all deposited on GitHub at https://github.com/rebeccamacpherson/dissertation_ch7_ARID1B_mods. DGRP 2.0 sequence data can be found at http://dgrp2.gnets.ncsu.edu. DGRP 3.0 sequence data will be available when the manuscript describing this resource is submitted for publication. All DGRP 2.0 and DGRP 3.0 flies are available from the Bloomington Drosophila Stock Center.

## Results

### *Drosophila Model of ARID1B-Related SSRIDD*

Previously, we established a Drosophila model for *ARID1B*-related SSRIDD using an *attp2 UAS-osa* TRiP RNAi line and the *Ubi156-GAL4* driver line (MacPherson *et al*. 2023). Here, we sought to validate an additional Drosophila model for *ARID1B*-related SSRIDD using an *attp40 UAS-osa* TRiP RNAi line crossed to the weak ubiquitous *Ubi156-GAL4* driver line. We first confirmed *osa* knockdown in *Ubi156>UAS-osa* males and females (~50% and ~60%, respectively) (Table S7.2). We then assessed

*Ubi156>UAS-osa* flies for changes in sleep and activity, as profound changes in sleep and activity and the mushroom body (a region of the Drosophila brain important for regulation of sleep (Joiner *et al*. 2006; Pitman *et al*. 2006; Sitaraman *et al*. 2015)) were previously reported in *ARID1B*-related SSRIDD flies (MacPherson *et al*. 2023). *ARID1B*-related SSRIDD patients also have hypotonia and central nervous system impairments (van der Sluijs *et al.* 2019; Schrier Vergano *et al.* 2021) that might be reflected in Drosophila *ARID1B*-related SSRIDD models as changes to sleep and activity. Similar to the decreased night sleep and increased activity reported for *ARID1B*-related SSRIDD models (MacPherson *et al*. 2023), we found decreases in night sleep ($p < 0.0018$ for males, $p < 0.0085$ for females) and day sleep ($p < 0.0049$ for males, $p < 0.0068$ for females) and increases in total activity ($p < 0.0031$ for males, $p < 0.0001$ for females) in both sexes, compared to the control, for our *Ubi156>osa ARID1B*-related SSRIDD models (Figure 7.1; Table S7.3).

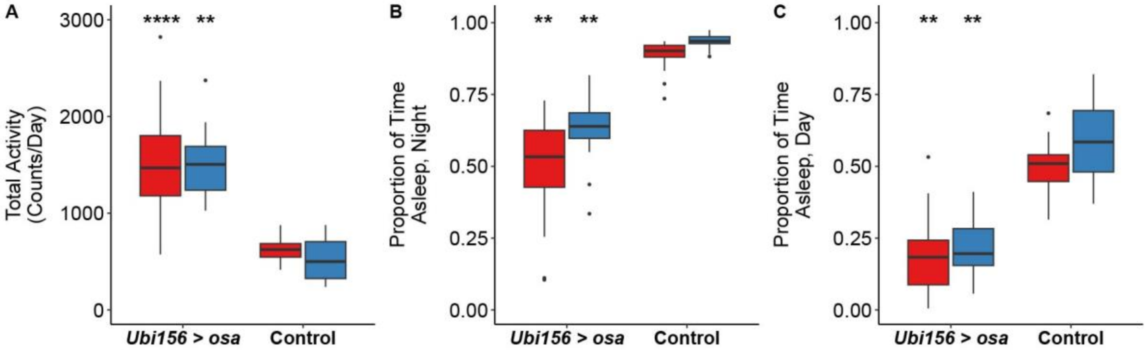


**Figure 7.1. Altered sleep and activity phenotypes in a Drosophila *ARID1B*-SSRIDD model.** Boxplots showing sleep and activity phenotypes for flies with *Ub156-GAL4*-mediated knockdown of *osa*, compared to the control. A: Total activity; B: proportion of

time sleep asleep at night; C: proportion of time spent asleep during the day. Females and males are shown in red and blue, respectively. See Supplementary Table 7.3 for statistical analyses. N=29-32 flies per sex per genotype. Asterisks (*) represent pairwise comparisons, sexes separately, of the knockdown line compared to the control. **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$

### *Genetic Modifiers for osa in the DGRP*

Having established a *Ubi156>(attp40)osa ARID1B*-related SSRIDD model, we then generated a Drosophila line that contained both the *Ubi156-GAL4* and *attp40 UAS-osa* elements (Figure S7.1). We maintained the resulting line (*CSB X; UAS-osa/CyO; Ubi-156-GAL4/TM3, Sb*) with balancer chromosomes due to the poor health of the strain. When crossed with the DGRP, this strain results in progeny with knockdown of *osa* across multiple DGRP genetic backgrounds. Thus, we can assess whether there are alleles segregating in different genetic backgrounds that alter changes in sleep and activity due to knockdown of *osa* (*i.e.* genetic modifiers).

We crossed virgin females from 10 maximally-homozygous, minimally-related, *Wolbachia*-negative DGRP lines to *Ubi156>osa* males and assessed sleep and activity phenotypes in *DGRP X; DGRP 2/TRiP 2; DGRP 3/Ubi-156-GAL4* (male) and *DGRP X/CSB X; DGRP 2/TRiP 2; DGRP 3/Ubi-156-GAL4* (female) progeny. Across all genetic backgrounds tested, flies with knockdown of *osa* were on average more active; they slept less during nighttime and daytime hours overall ($p > 0.0001$; Table S7.4). Changes to night sleep, day sleep, and daytime sleep bout length as a result of *osa* knockdown were all dependent on genetic background ($p < 0.0001$ for all phenotypes and sexes; Figure 7.2; Table S7.4). Changes to total activity as a result of *osa* knockdown were dependent upon genetic background for males only ($p < 0.0001$; Figure 7.2; Table S7.4). Interestingly, although genetic background changes the degree to which *osa* knockdown impacts daytime sleep bout length, knockdown of *osa* alone did not significantly change daytime sleep bout length compared to the control ($p > 0.05$, Table S7.4). These background dependent effects of *osa* knockdown on sleep and activity phenotypes suggest that genetic modifiers for *osa* segregate in the DGRP.

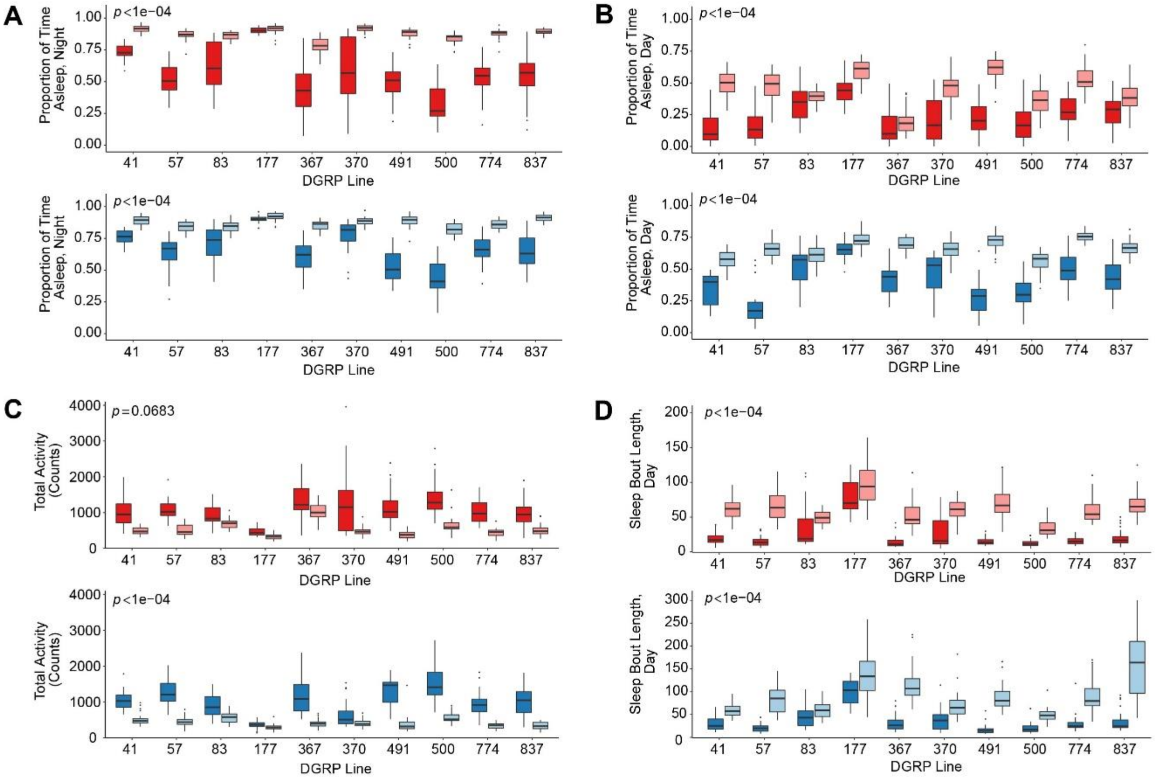


**Figure 7.2. Background-dependent sleep and activity changes in Drosophila *ARID1B*-SSRIDD models.** Boxplots showing sleep and activity phenotypes for flies with *Ub156*-

*GAL4*-mediated knockdown of *osa* across different DGRP lines. A: proportion of time spent asleep at night; B: proportion of time spent asleep during the day; C: total activity; D: sleep bout length during the day. Females and males are shown in red and blue, respectively. Flies with knockdown of *osa* and flies without knockdown of *osa* are represented by darker and lighter shades, respectively. See Supplementary Table 7.4 for statistical analyses. N=12-32 flies per sex per genotype per DGRP line. *p* values shown represent the effect of the interaction between genotype (*osa* knockdown versus control) and line (DGRP line). DGRP line numbers refer to DGRP 2.0.

### *Assessment of Wing Vein Phenotypes*

Although *CSB X; UAS-osa/CyO; Ubi-156-GAL4/TM3, Sb* flies are suitable to identify genetic modifiers for *osa*, these flies are too sickly to be reared in sufficient quantities for a large-scale genome wide association analysis. Thus, we took advantage of a tissue-specific model for *ARID1B*-related SSRIDD to avoid the deleterious fitness effects associated with decreased expression of *osa*.

We generated a wing-specific *ARID1B*-related SSRIDD Drosophila model using the crossing scheme outlined in Supplementary Figure S7.2 with the *MS1096-GAL4* driver, which drives expression of *UAS* in the dorsal wing disc. *MS1096>osa* flies have varying degrees of impaired wing vein formation, with incomplete longitudinal veins and/or crossveins (Figure 7.3). We devised a scoring system to reflect these wing vein changes, scoring each wing between zero and six, based on the total number of visibly intact longitudinal veins (L2, L3, L4, L5) and/or crossveins (anterior, posterior) (Figure 7.3). For the GWA screen, we crossed *MS1096>osa* females to DGRP males from 392 DGRP lines and scored the right wings from 15 randomly selected progeny of each cross, for males (*MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa*) and females (*DGRP X/MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa*) separately. Due to the presence of the *GAL4* driver on the *X* chromosome, we also performed analyses for males and females separately.

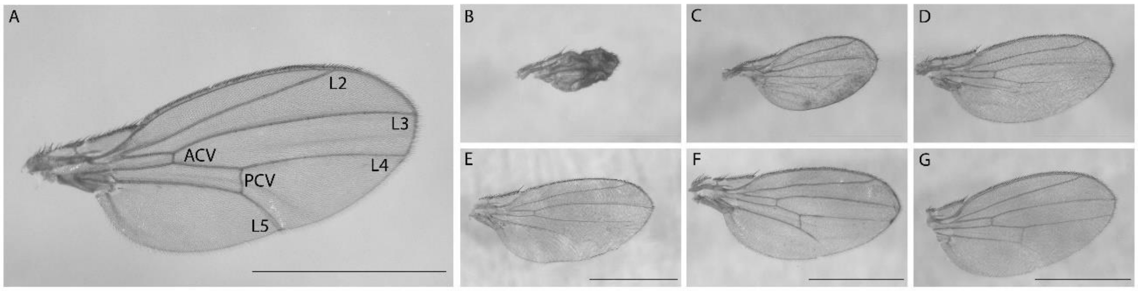


**Figure 7.3. Examples of wings with altered veins.** Images of right wings from the *MS1096>osa* DGRP screen. A: Wing from a control male (*DGRP2.0; DGRP321 X; DGRP321 2/TRiP 2*; *DGRP321 3/UAS-osa*) showing intact Longitudinal vein 2 (L2), Longitudinal vein 3 (L3), Longitudinal vein 4 (L4), Longitudinal vein 5 (L5), Anterior crossvein (ACV), and posterior crossvein (PCV) for a score of 6; B: wing from a male fly (*DGRP2.0; MS1096-GAL4; DGRP409 2/TRiP 2*; *DGRP409 3/UAS-osa*) showing a crumpled phenotype with no visibly intact wing veins for a score of 0; C: wing from a male fly (*DGRP2.0; MS1096-GAL4; DGRP41 2/TRiP 2*; *DGRP41 3/UAS-osa*) showing an intact L2 for a score of 1; D: wing from a female fly (*DGRP3.0; RP_0493 X/MS1096-GAL4; RP_0493 2/TRiP 2*; *RP_0493 3/UAS-osa*) showing an intact L2 and ACV for a

score of 2; E: wing from a female fly (*DGRP3.0; RP_0491 X/MS1096-GAL4; RP_0491 2/TRiP 2*; *RP_0491 3/UAS-osa*) showing an intact L2, L3, and ACV for a score of 3; F: wing from a female fly (*DGRP2.0; DGRP367 X/MS1096-GAL4; DGRP367 2/TRiP 2*; DGRP367 *3/UAS-osa*) showing an intact L2, L3, ACV, and PCV for a score of 4; G: wing from a female fly (*DGRP2.0; DGRP83/MS1096-GAL4; DGRP83 2/TRiP 2*; *DGRP83 3/UAS-osa*)showing an intact L2, L3, L4, ACV, and PCV for a score of 5. Not all possible intact wing vein and score combinations are shown. Scale bar represents 1mm.

We ranked line means (Table S7.5) for males and females separately and selected three subsets of lines for each sex for which we would also score wings from 15 relevant control flies (*DGRP X/FM6, B; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa* (females) and *FM6, B; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa* (males)): the lines with the 40 greatest and 40 smallest line mean values, as well as 20 lines on either side of the population median for a total of 120 lines per sex. We thus aimed to score wings from about one-third of control flies collected. If lines with equal line means were on either side of a threshold, wings from all lines with that line mean were scored (*e.g.* if the 39th-42nd lines had the same line means, then wings from 42 lines were scored for that sex and threshold, as opposed to 40). This resulted in wings scored from 162/392 and 164/392 DGRP lines for males and females, respectively (Table S7.6). Out of the 4746 control fly wings scored, 99.8% (4738/4746) scored a “6” and all others (8/4746) scored a “5” (Table S7.6). Each of the 8 control wings with a score of “5” were from different genetic backgrounds. Given the rarity of control wings without a 6/6 score, we did not include scores from control flies in our analyses of variance, as wing vein score does not significantly vary across the DGRP lines tested.

The distribution of wing vein score line means for knockdown flies is shown in Figure 7.4, ranked in ascending order, sexes separately. At least one wing scored between 0-6 (inclusive) for each sex. Wing vein score in flies with knockdown of *osa does* vary with genetic background ($p < 0.0001$) and sex (p < 0.0001; Table S7.7). On average, females had a higher line mean (3.14) than males (2.74). Scores from male and female flies are moderately positively correlated with one another (Pearson's $r = 0.4960$; Figure S7.3). There is variation in wing vein score, as the broad sense heritability ($H^2$) is 0.28 for both sexes together, and 0.24 and 0.32 for males and females, respectively (Table S7.7). Heritability based on line means was higher, at 0.82 and 0.88 for males and females, respectively (Table S7.7).

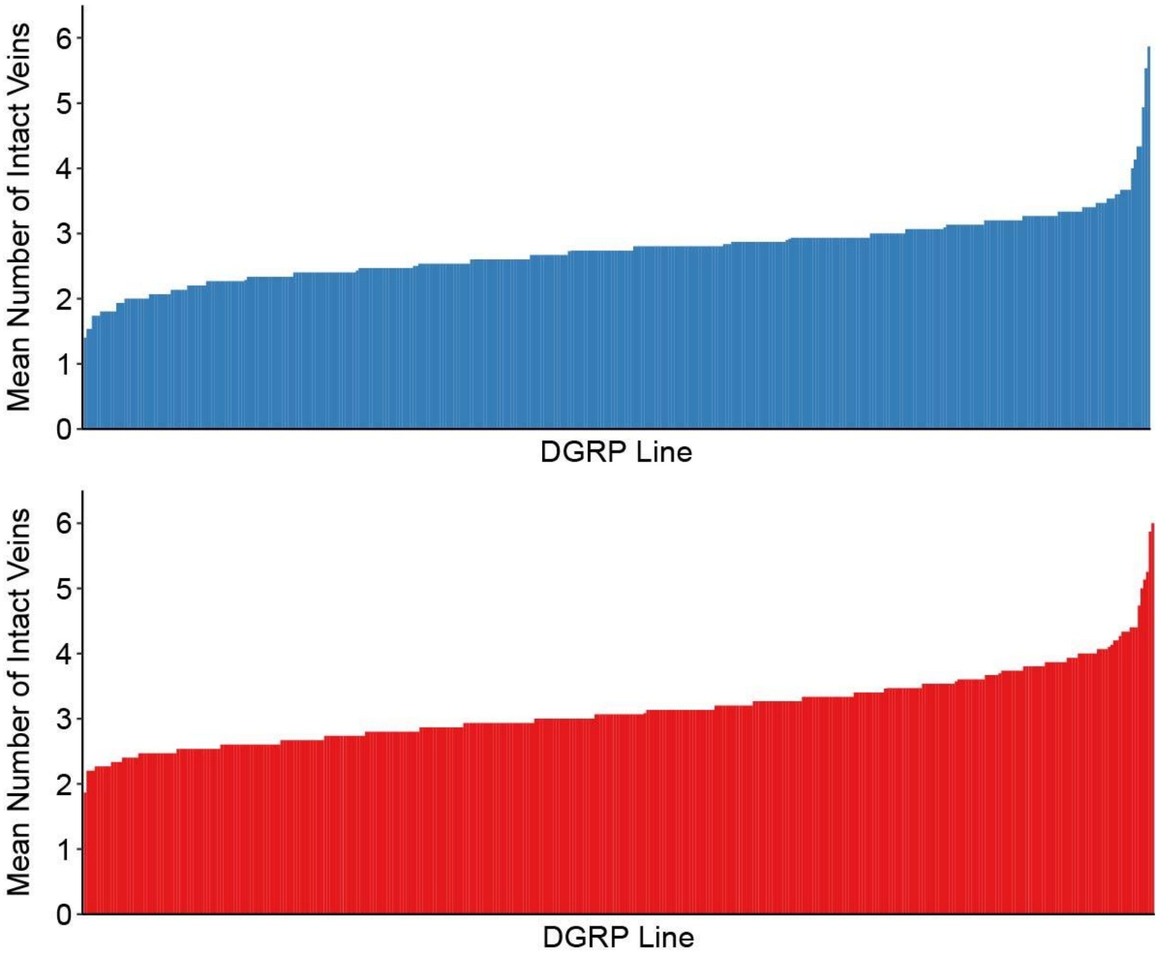


**Figure 7.4. Line means for wing vein scores.** Bar graphs showing line means of intact wing vein scores for A: females (*DGRP X/MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa*) and B: males (*MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa*). Line means are shown in ascending order, sexes separately. Graphs represent 392 DGRP lines with roughly 15 flies per sex per line.

## Discussion

We used the *UAS-GAL4* system to model *ARID1B*-related SSRIDD in *D. melanogaster* and present evidence that variants modifying *osa* (the Drosophila ortholog of *ARID1B*) are segregating in the DGRP. In agreement with a previous RNAi-mediated *ARID1B*-related SSRIDD fly model, flies with weak ubiquitous knockdown of *osa* show decreased night sleep and increased activity. By crossing *Ubi156>osa* flies to the DGRP, we observe that *Ubi156>osa*-mediated changes in sleep and activity are dependent on genetic background. However, given the decreased fitness of flies with ubiquitous knockdown of *osa*, we turned to a tissue-specific model of *ARID1B*-related SSRIDD for further experimentation, using the wing-specific GAL4 driver *MS1096*. *MS1096>osa* flies have impaired wing veination. Thus, we crossed the wing-specific *ARID1B*-SSRIDD fly model to 392 DGRP lines and found a high degree of genetic variation for the number of intact wing veins in resulting progeny. These data will be used to perform GWA analyses to identify candidate genetic modifiers for *osa*. Validated candidate genetic modifiers may translate to human populations and provide a genetic explanation for the phenotypic variation associated with *ARID1B* variants.

Although all DGRP 3.0 lines have been generated (T. Mackay, personal communication, March 2023), resulting DNA sequencing data are currently under analysis. Additionally, fully genotyped (Huang *et al*. 2014) DGRP 2.0 lines have been re-sequenced and DNA sequencing data from DGRP 2.0 and 3.0 are being genotyped together (T. Mackay, personal communication, Apr 2023). A new GWA analysis pipeline for DGRP data is under development, which will account for genetic relatedness, *Wolbachia* infection status, and effects of inversions (T. Mackay, personal communication, Apr 2023).

We will use this new pipeline to perform GWA analyses for *MS1096>osa*-mediated wing vein scores, performing association tests for DNA sequence variants present in the 392 tested DGRP lines with a minor allele frequency (MAF) greater than 0.05. The exact number of DNA sequence variants with MAF>0.05 in the new DGRP lines is currently unknown, although we expect at least 2.5 million common (MAF > 0.05) DNA sequence variants (~2.5 million common variants were observed in the 205 lines of the DGRP 2.0 (Huang *et al*. 2014)). We will first perform GWA analyses using female phenotype data on chromosomes *X*, *2*, and *3* and male data on chromosomes *2* and *3*. We can only identify associated sequence variants using male data on chromosomes *2* and *3*, as the *MS1096-GAL4* driver is the only *X* chromosome present in the phenotyped males (Drosophila males only have one *X* chromosome, which they inherit from their female parent). However, given the moderate correlation of male and female wing vein scores, we could also perform an analysis using data from both sexes. Based on previous GWA studies using the DGRP 2.0 (Chow *et al.* 2016; Anholt and Mackay 2017; Lavoy *et al.* 2018; Mackay and Huang 2018; Talsness *et al.* 2020), we are confident that we will identify polymorphisms associated with variation in wing vein score, though it is likely that only a few will surpass the strict Bonferroni significance threshold. We may also use a lower significance threshold; studies using the DGRP 2.0 typically use a threshold of $p < 10^{-5}$ (Mackay and Huang 2018). We will not include a control population in the GWA analysis, as almost all control flies examined (4738/4746; 99.8%) had a wild-type wing vein phenotype and therefore do not show variation in our phenotype of interest.

Following identification of associated polymorphisms, we will perform functional analyses of genes implicated by GWA analyses. Following convention, variants falling in or near (within 1kb) coding genes will be attributed to that gene. Intragenic variants are more challenging to interpret and likely fall within transcription factor binding domains, areas tied to chromatin configuration, and/or microproteins, and often cannot be linked to a specific gene.

We will validate associated genes through DGRP out-of-sample testing. We will select variants with the strongest associations in both sexes (whether on chromosomes *2* and *3* from sex-specific GWA analyses, or on chromosomes *X, 2, or 3* from the GWAS containing data from both sexes) for validation. The exact number of variants will be dependent upon GWA results. For each candidate variant, we will randomly select 20 DGRP lines with the primary allele and 20 DGRP lines with the alternate allele from DGRP lines not previously tested. We will cross males from each of these DGRP lines to *MS1096>osa* virgin females and score wing veins from at least 30 *MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa* males and 30 *DGRP X/MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa* females. We will then perform a *t*-test to determine if the wing vein scores from flies with the alternate allele differ from those with the primary allele. Statistically different wing vein scores indicate that the variant of interest is a validated candidate genetic modifier for *osa*.

We are confident that we will identify and validate candidate genetic modifiers using this approach. Previously, subsets of the DGRP 2.0 were crossed to a focal mutation and progeny analyzed to identify SNPs associated with Drosophila traits of interest (He *et al.* 2016; Palu *et al.* 2019; Özsoy *et al.* 2021). Others have used this approach to identify and validate potential candidate genetic modifiers for human disorders (Chow *et al.* 2016; Lavoy *et al.* 2018; Talsness *et al.* 2020). Furthermore, dozens of studies have identified modifiers in the DGRP without use of a focal mutation (Anholt and Mackay 2018). One of the primary differences of our study is the use of additional DGRP lines, which should provide increased statistical power and therefore permit detection of a larger number of overall variants and/or variants with smaller effect sizes. The DGRP 3.0 also provides a larger number of genetic backgrounds for out-of-sample testing, thus providing additional statistical power at the validation stage as well.

The presence of genetic modifiers for *osa* in the DGRP suggests that genetic modifiers at least partially explain the wide phenotypic spectrum associated with variants in *ARID1B.* Once candidate genetic variants are identified, it will be intriguing to hypothesize mechanistic and functional links between the associated genes and *ARID1B.* Modifier genes could be involved in transcriptional regulation, chromatin modification, neurological development, and/or an underappreciated or novel role for *ARID1B. osa* is also orthologous to the human gene *ARID1A.* Although *ARID1A* also encodes a core structural subunit of the mammalian SWI/SNF complex, it is less commonly associated with SSRIDDs and is more commonly associated with cancers than *ARID1B.* Candidate genetic modifiers for *ARID1B* identified using *osa* knockdown may also be candidate genetic modifiers for *ARID1A*. Regardless, the presence of genetic modifiers supports a spectrum-based approach to *ARID1B*-related disorders, including autism spectrum disorder and SSRIDDs.

## Acknowledgements

We thank Dr. Nestor O. Nazario-Yepiz for his assistance with crossing schemes and the Clemson Center for Human Genetics Bioinformatics and Statistics Laboratory for their computational help. We thank the TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing transgenic RNAi fly stocks used in this study.

This work was funded by NIH grants R01 GM128974 to TFCM and RRHA and F31 HD106719 to RAM.

## References

American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders. Fifth edition ed. Washington, DC: American Psychiatric Association Publishing.

Anholt RRH, Mackay TFC. The road less traveled: From genotype to phenotype in flies and humans. *Mamm Genome.* 2018. 29(1-2):5-23

Bögershausen N, Wollnik B. Mutational landscapes and phenotypic spectrum of SWI/SNF-related intellectual disability disorders. *Front Mol Neurosci.* 2018. 11:252.

Cenik BK, Shilatifard A. COMPASS and SWI/SNF complexes in development and disease. *Nat Rev. Genet.* 2021. 22(1):38-58.

Centore RC, Sandoval GJ, Soares LMM, Kadoch C, Chan HM. Mammalian SWI/SNF chromatin remodeling complexes: Emerging mechanisms and therapeutic strategies. *Trends Genet.* 2020. 36(12):936-50.

Chow CY, Kelsey KJP, Wolfner MF, Clark AG. Candidate genetic modifiers of retinitis pigmentosa identified by exploiting natural variation in Drosophila. *Hum Mol Genet.* 2016. 25(4):651-9.

Chubak MC, Nixon KCJ, Stone MH, Raun N, Rice SL, Sarikahya M, Jones SG, Lyons TA, Jakub TE, Mainland RLM, *et al*. Individual components of the SWI/SNF chromatin remodelling complex have distinct roles in memory neurons of the Drosophila mushroom body. *Dis Model Mech.* 2019. 12(3):dmm037325.

Cichewicz K, Hirsh J. ShinyR-DAM: A program analyzing Drosophila activity, sleep and circadian rhythms. *Commun Biol.* 2018. 1(1):25.

Cukier HN, Perez AM, Collins AL, Zhou Z, Zoghbi HY, Botas J. Genetic modifiers of MeCP2 function in Drosophila. *PLoS Genet.* 2008. 4(9):e1000179.
Ellegood J, Petkova SP, Kinman A, Qiu LR, Adhikari A, Wade AA, Fernandes D, Lindenmaier Z, Creighton A, Nutter LMJ, *et al*. Neuroanatomy and behavior in mice with a haploinsufficiency of AT-rich interactive domain 1B (ARID1B) throughout development. *Mol Autism.* 2021. 12(1):25.

Garlapow M, Huang W, Yarboro M, Peterson K, Mackay T. Quantitative genetics of food intake in *Drosophila melanogaster*: e0138129. *PloS One.* 2015. 10(9).

Grozeva D, Carss K, Spasic-Boskovic O, Tejada M, Gecz J, Shaw M, Corbett M, Haan E, Thompson E, Friend K, *et al*. Targeted next-generation sequencing analysis of 1,000 individuals with intellectual disability. *Hum Mutat.* 2015. 36(12):1197-204.

He S, Wu Z, Tian Y, Yu Z, Yu J, Wang X, Li J, Liu B, Xu Y. Structure of nucleosome-bound human BAF complex. *Science.* 2020. 367(6480):875-81.

He X, Zhou S, St Armour GE, Mackay TF, Anholt RR. Epistatic partners of neurogenic genes modulate Drosophila olfactory behavior. *Genes Brain Behav.* 2016. 15(2):280-90.

Hoyer J, Ekici A, Endele S, Popp B, Zweier C, Wiesener A, Wohlleber E, Dufke A, Rossier E, Petsch C, *et al*. Haploinsufficiency of *ARID1B*, a member of the SWI/SNF-A chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am J Hum Genet.* 2012. 90(3):565-72.

Hu X, Li M, Hao X, Lu Y, Zhang L, Wu G. The *osa*-containing SWI/SNF chromatin-remodeling complex is required in the germline differentiation niche for germline stem cell progeny differentiation. *Genes.* 2021. 12(3):363.

Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF, *et al*. Natural variation in genome architecture among 205 *Drosophila melanogaster* genetic reference panel lines. *Genome Res.* 2014. 24(7):1193-208.

Joiner WJ, Crocker A, White BH, Sehgal A. Sleep in Drosophila is regulated by adult mushroom bodies. *Nature.* 2006. 441(7094):757-60.

Jung E, Moffat JJ, Liu J, Dravid SM, Gurumurthy CB, Kim W. *Arid1b* haploinsufficiency disrupts cortical interneuron development and mouse behavior. *Nat Neurosci.* 2017. 20(12):1694-707.

Kadoch C, Crabtree GR. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv.* 2015. 1(5):e1500447.

Kadoch C, Williams RT, Calarco JP, Miller EL, Weber CM, Braun SMG, Pulice JL, Chory EJ, Crabtree GR. Dynamics of BAF-polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat Genet.* 2017. 49(2):213-22.

Lavoy S, Chittoor-Vinod V, Chow CY, Martin I. Genetic modifiers of neurodegeneration in a Drosophila model of Parkinson's disease. *Genetics.* 2018. 209(4):1345-56.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Δ Δ C(T)) method. *Methods.* 2001. 25(4):402-8.

Mackay TFC, Huang W. Charting the genotype–phenotype map: Lessons from the *Drosophila melanogaster* genetic reference panel. *Wiley Interdiscip Rev Dev Biol.* 2018. 7(1):e289.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, *et al*. The *Drosophila melanogaster* genetic reference panel. *Nature.* 2012. 482(7384):173-8.

MacPherson RA, Shankar V, Anholt R, Mackay T. Genetic and genomic analyses of *Drosophila melanogaster* models of chromatin modification disorders. *Genetics.* 2023. iyad061

Michel BC, D'Avino AR, Cassel SH, Mashtalir N, McKenzie ZM, McBride MJ, Valencia AM, Zhou Q, Bocker M, Soares LMM, *et al*. A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol.* 2018. 20(12):1410-20.

Özsoy ED, Yılmaz M, Patlar B, Emecen G, Durmaz E, Magwire MM, Zhou S, Huang W, Anholt RRH, Mackay TFC. Epistasis for head morphology in *Drosophila melanogaster*. *G3 (Bethesda).* 2021. 11(10).

Palu RAS, Ong E, Stevens K, Chung S, Owings KG, Goodman AG, Chow CY. Natural genetic variation screen in Drosophila identifies Wnt signaling, mitochondrial metabolism, and redox homeostasis genes as modifiers of apoptosis. *G3 (Bethesda).* 2019. 9(12):3995-4005.

Perkins LA, Holderbaum L, Tao R, Hu Y, Sopko R, McCall K, Yang-Zhou D, Flockhart I, Binari R, Shim H, *et al*. The transgenic RNAi project at Harvard medical school: Resources and validation. *Genetics.* 2015. 201(3):843-52.

Pitman J, McGill J, Keegan K, Allada R. A dynamic role for the mushroom bodies in promoting sleep in Drosophila. *Nature.* 2006. 441(7094):753-6.
Satterstrom FK, Breen MS, Grove J, Klei L, Xu X, Norman U, Brand H, Schwartz G, Barbosa M, Bybjerg-Grauholm J, *et al*. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell.* 2020. 180(3):568,584.e23.

Schrier Vergano S, Santen G, Wieczorek D, Wollnik B, Matsumo N, Deardorff MA. 2013 [Updated 2021]. Coffin-Siris syndrome. In: GeneReviews (R). Adam MP, Everman DB, Mirzaa GM,*et al*., editors. Seattle, WA: University of Washington, Seattle.

Shibutani M, Horii T, Shoji H, Morita S, Kimura M, Terawaki N, Miyakawa T, Hatada I. *Arid1b* haploinsufficiency causes abnormal brain gene expression and autism-related behaviors in mice. *Int J Mol Sci.* 2017. 18(9):1872.

Sitaraman D, Aso Y, Rubin GM, Nitabach MN. Control of sleep by dopaminergic inputs to the Drosophila mushroom body. *Front Neural Circuits.* 2015. 9(73).

Sokpor G, Xie Y, Rosenbusch J, Tuoc T. Chromatin remodeling BAF (SWI/SNF) complexes in neural development and disorders. *Front Mol Neurosci.* 2017. 10:243.

Talsness DM, Owings KG, Coelho E, Mercenne G, Pleinis JM, Partha R, Hope KA, Zuberi AR, Clark NL, Lutz CM, *et al.* A Drosophila screen identifies NKCC1 as a modifier of NGLY1 deficiency. *eLife.* 2020. 9.

Tian Y, Smith-Bolton RK. Regulation of growth and cell fate during tissue regeneration by the two SWI/SNF chromatin-remodeling complexes of Drosophila. *Genetics.* 2021. 217(1):1-16.

van der Sluijs PJ, Jansen S, Vergano SA, Adachi-Fukuda M, Alanay Y, AlKindy A, Baban A, Bayat A, Beck-Wödl S, Berry K, *et al*. The ARID1B spectrum in 143 patients: From nonsyndromic intellectual disability to Coffin–Siris syndrome. *Genet Med.* 2019. 21(6):1295-307.

Vasko A, Drivas TG, Schrier Vergano SA. Genotype-phenotype correlations in 208 individuals with Coffin-Siris syndrome. *Genes.* 2021. 12(6):937.

Vergano SA, van der Sluijs PJ, Santen G. 2019. ARID1B-related disorder. In: GeneReviews (R). Adam MP, Mirzaa GM, Pagon RA,*et al*., editors. Seattle, WA: University of Washington, Seattle.

Yuen RKC, Merico D, Bookman M, Howe JL, Thiruvahinapuram B, Patel RV, Whitney J, Deflaux N, Bingham J, Wang Z, *et al*. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci.* 2017. 20(4):602-11.

Zhou S, Morgante F, Geisz MS, Ma J, Anholt RRH, Mackay TFC. Systems genetics of the Drosophila metabolome. *Genome Res.* 2020. 30(3):392-405.

Zirin J, Hu Y, Liu L, Yang-Zhou D, Colbeth R, Yan D, Ewen-Campen B, Tao R, Vogt E, VanNest S*, et al*. Large-scale transgenic Drosophila resource collections for loss- and gain-of-function studies. *Genetics.* 2020. 214(4):755-67.

# CHAPTER EIGHT

# DISCUSSION AND CONCLUSIONS

I have investigated the genetic underpinnings of neurodevelopmental disorders FASD, CdLS, and SSRIDDs using the *D. melanogaster* model system. Embracing variation across individuals within a single genetic background, I developed a novel assay quantifying time-to-sedation from exposure to ethanol vapors in Chapter II. I used this assay in Chapters III and V, which present changes in the transcriptome alongside changes in behavior as a result of introducing stressors on Drosophila, whether environmental (alcohol) or genetic (gene deletion). Chapter III shows sex-specific changes in neuronal and glial cell populations in brain tissue, as well as decreased viability, increased time to sedation, and changes in sleep due to developmental ethanol exposure. This indicates that developmental ethanol exposure has widespread effects throughout the brain and impacts adult behaviors. Chapter IV demonstrates that changes in sleep and the transcriptome resulting from developmental ethanol exposure vary not only with sex, but also by genetic background. Chapter IV also identified many background-dependent genetic networks important for response to developmental ethanol exposure, including a network of largely small-nucleolar RNAs. Chapter V investigates this snoRNA network further by providing the first characterization of *Uhg4*, which is moderately correlated in expression with these snoRNAs and is critical for reproduction and stress-response. Transcriptomic data across our Drosophila FASD models indicates that the genetic architecture of response to developmental ethanol exposure is complex, dependent on sex and/or genetic background, and persistent into adulthood. Chapter VI focuses on Drosophila models of other neurodevelopmental disorders, including SSRIDDs and CdLS, yet the major takeaways remain the same. Decreased expression of individual members of the cohesin and/or BAP complex reveals extensive sex- and gene-dependent changes in the transcriptome, sleep and activity, brain structure, and sensorimotor integration. Chapter VII establishes another Drosophila model for *ARID1B*-SSRIDD and presents evidence for genetic modifiers of *osa*, the Drosophila ortholog of *ARID1B*. Chapter VI identifies genes that may be candidate genetic modifiers for *SMARCB1*-related SSRIDD and Chapter VII lays the groundwork for the identification of candidiate genetic modifiers for *ARID1B*-related SSRIDD. These findings indicate that even within a single neurodevelopmental disorder, the genetic architecture is complex, and the role of genetic background cannot be ignored.

Many assays for quantifying Drosophila sedation do not account for individual variation; large numbers of Drosophila are assessed in groups and a group average or a proportion serves as the focal data point (Singh and Heberlein 2000). Chapter II provides the first high-throughput assay to measure sedation time of individual Drosophila, capturing variation that is typically ignored. This variation provides information about within-group variation that can differentiate two populations (*i.e.* genetic backgrounds) from one another that might appear similar based on group mean alone.

The role of genetic background is relevant to all chapters of this dissertation. The single cell RNA sequencing studies in Chapter III were only performed on one genetic background, yet response to ethanol exposure is background-dependent (Chapter IV in this dissertation; Morozova *et al.* 2015; 2018). McClure *et al.* (2011) and Bonilla *et al.* (2021) also show that effects of ethanol exposure in Drosophila vary with time point of exposure and persist across generations. Therefore, additional single cell RNA sequencing studies on brain tissue from different genetic backgrounds, developmental stages, and generations could reveal how pervasive the changes identified in Chapter III are across genetic backgrounds and timepoints. These future directions emphasize the limitations of results from Chapter III; although conservation of ethanol metabolism suggests the broad findings are likely generalizable, more specific conclusions most likely are tied to the exact experimental setup. Future work in this area could also take advantage of novel single cell technologies that can capture the epigenome and transcriptome in parallel. This future direction would be of particular interest, given the ability of alcohol to modify the epigenome (Liu, Balarman *et al.* 2009; Wallen *et al.* 2021).

In contrast with Chapter III, transcript data from Chapters IV and VI are from whole-body Drosophila tissue; assessment of specific tissues related to neurological development (*i.e.* brain or ventral nerve cord) may provide more relevant insights for these neurodevelopmental disorders. Important signals from individual tissues may be masked by conglomeration of all tissue types together, just as signals from individual cell types may be masked in a heterogeneous tissue type. Future studies should build upon the work described herein and examine differential expression across individual tissues and multiple time points.

Differentially expressed genes identified in Chapters III, IV, V and VI include a large number of non-coding RNAs. Although noncoding RNAs are generally acknowledged to be important, relatively few studies have attempted to individually characterize functional roles for noncoding RNAs. The work discussed in Chapter V is the first characterization of the lncRNA *Uhg4*. Additionally, the generated *Uhg4* deletion lines are the first known Drosophila lines that feature a deletion of *Uhg4* alone. Although snoRNA host genes are difficult to delete with CRISPR-Cas9 due to the presence of intronic snoRNAs, Chapters IV and V indicate that expression of host genes is likely independent of the expression of the snoRNAs which they host. Combined with the fitness effects shown in Chapter V, these results emphasize that snoRNA host genes have roles independent of the snoRNAs which they host. Thus, a method for perturbing individual snoRNAs without impacting nearby snoRNAs or the host gene would be beneficial to provide greater clarity as to the role of *Uhg4* in Drosophila but also the roles for snoRNAs and their host genes.

In addition to highlighting the individual importance of snoRNAs and their host genes, this dissertation also emphasizes the importance of noncoding RNAs more broadly. All transcriptomic data generated herein identify critical genetic networks and hub genes that center around noncoding RNAs, but a role or function for many of these RNAs in the organism remains unknown. Although additional research on the thousands of uncharacterized noncoding RNAs is needed, noncoding RNAs are still challenging to study. Some of the barriers to working with noncoding RNAs are that they may be difficult to genetically manipulate, may not have publicly available reagents, and may be considered high risk or a potential dead-end. snoRNAs are important for stress response but are difficult to perturb without affecting other genes. How much functional information is missed due to the lack of knowledge of noncoding RNAs? Perhaps new CRISPR technologies with alternative PAM sites and new transcript sequencing technologies will facilitate more in-depth study of noncoding RNAs. This dissertation affirms that noncoding RNAs are important and are worthy of future research.

The role of *Uhg4* in response to multiple stressors (ethanol and temperatures), and the overlap in changes in gene expression in the Drosophila brain upon exposure to two “stressful” different foreign substsances (ethanol and cocaine) suggests that there is overlap across different stress responses. Thus, although this dissertation focuses on ethanol response, future studies should investigate the generalizability of ethanol response to other teratogens (in the case of developmental ethanol exposure) and/or other drugs of abuse. The findings from Chapter V do have limited direct translational potential, as *Uhg4* does not have a human ortholog (Hu *et al.* 2011). However, *Uhg4* is coregulated with many genes that do have human orthologs, and although snoRNA sequences are not well-conserved, snoRNA structure and roles in ribosomal RNA modification are conserved (Czekay and Kothe 2021). This suggests that despite a lack of sequence similarity, snoRNA gene function and interactions may be conserved.

Drosophila are capable of experience-dependent changes in behavior, which has been used as a model for intellectual disability (Coll-Tane *et al.* 2019). Intellectual disability is a common clinical finding for all neurodevelopmental disorders studied in this dissertation yet assessment of this phenotype in our Drosophila models was outside the scope of this work. Although some groups have examined learning and memory in Drosophila models of SSRIDDS and CdLS (Wu *et al.* 2015; Chubak *et al.* 2019), these studies either did not account for genetic background or used cell-type specific gene knockdown in their Drosophila models. Thus, learning and memory assays would be an interesting future direction for numerous projects included in this dissertation.

In addition to learning and memory, future studies should also consider how chromatin accessibility differs across SSRIDD and/or CdLS Drosophila models. Chapter VI focuses on the transcriptome and potential genetic modifiers, but given the role of the BAP and cohesin complexes in chromatin remodeling, the assay for transposase-accessible chromatin with sequencing (ATAC-seq) or single-cell multiome sequencing remains an intriguing direction.

Although Drosophila have been used to model one or more subtypes of SSRIDDs and CdLS (Wu *et al.* 2015; Chubak *et al.* 2019), Chapter VI establishes the first SSRIDD and

CdLS Drosophila models with ubiquitous, as opposed to tissue-specific, gene knockdown. For some SSRIDD and CdLS subtypes, Chapter VI establishes the first animal models of any kind. Rarer subtypes of CdLS such as *SMC3*-associated CdLS have not been studied as extensively as the more common *NIPBL*-associated CdLS, thus knowledge gaps remain for these rarer subtypes. The validation of these SSRIDD and CdLS Drosophila models should provide a jumping off point for future work. There are, however, several other subtypes that were not studied in this dissertation. Some of these SSRIDD- and CdLS-associated genes were not included because the gene of interest was not associated with the disorder at the time of project initiation, while others were not included because ubiquitous knockdown of the gene(s) were lethal in Drosophila. The reasons why ubiquitous knockdown of only *some* SSRIDD associated genes resulted in lethality remains unknown. Our SSRIDD and CdLS models were only representative of moderate gene knockdown, which is not representative of all variants reported for SSRIDD- and CdLS-associated genes. Future studies could develop SSRIDD and/or CdLS models with specific pathogenic patient variants, though development of patient-specific Drosophila models proved exceptionally challenging for our laboratory. Findings from patient-specific models should provide a higher degree of translatability.

Most Drosophila human disease models approach a disorder one gene or one associated variant at a time (Bellen *et al.* 2019). Although this is a valuable method for mechanistic studies or associating a variant with a disease phenotype, such a narrow focus has limited generalizability to other disease subtypes or genetic backgrounds. Larger-scale studies comparing multiple genotypes, such as those in Chapter VI, are limited by the resources required for CRISPR-based approach that preserves genotype, in addition to the challenges associated with placing and maintaining a deleterious mutation in a live organism. Chapters VI and VII use a *UAS-GAL4* RNAi-based approach with a weak *GAL4* driver to allow for high-throughput generation of primarily viable Drosophila models. This allows for comparisons across subtypes and across disorders, a particularly valuable approach as SSRIDDs and CdLS shift towards spectrum-based classifications (Bogershausen and Wollnik 2018; Kline *et al.* 2018).

Prior to the work described in the dissertation, transcriptome data from SSRIDD patients or animal models did not exist, with the exception of a single *ARID1B*-SSRIDD patient lymphocyte cell line (Vasileiou *et al.* 2015). For CdLS, transcriptional data primarily exists for specific tissues or cell types of *NIPBL*-CdLS (Kawauchi *et al.* 2009; Liu, Zhang *et al.* 2009; Revenkova *et al.* 2009; Wu *et al* 2015; Mills *et al.* 2018; Weiss *et al.* 2018). Chapter VI provides transcriptome-based insight into the pathogenesis of multiple subtypes of SSRIDDs and CdLS. Chapter VI revealed that a single focal genetic perturbation (*i.e.* gene knockdown) results in thousands of differentially expressed genes. This is the first whole-transcriptome dataset comparing subtypes of SSRIDDs or CdLS in humans or in animal models. The widespread changes observed make it challenging to pinpoint causal mechanisms of disease pathogenesis and instead suggest broad implications in development. Co-regulated genes from Chapter VI provide starting points for medical and basic researchers alike in understanding how changes in chromatin remodelers leads to disease.

Due to the rare nature of SSRIDDs, there have been no GWA studies performed using human patient data. Chapter VI contains the groundwork for what would be the first GWA study performed on SSRIDD animal models. Apart from the intended GWA study, Chapter VI also suggests the presence of genetic modifiers for *osa*, and therefore modifiers for *ARID1B* are probable. Combined with previous successes using the DGRP to identify trait-modifiying varaints (Mackay and Huang 2018), it is likely that at least some GWA hits will be validated. Future work will focus on identifying and validating these candidate modifiers. *ARID1B*-associated SSRIDD is one of only a handful of human disorders with formal scientific evidence generated using Drosophila animal models implicating genetic modifiers in disease pathogenesis (Buff *et al.* 2007; Branco *et al.* 2008; Cukier *et al.* 2008; Zhan *et al.* 2013; Chow *et al.* 2016; Lavoy *et al.* 2018; Palu *et al.* 2019; Talsness *et al.* 2020).

The presence of genetic modifiers may contribute to the lack of clear genotype-phenotype correlation for SSRIDDs. Drosophila models can be used to identify potential modifiers, though even validated modifiers in Drosophila may not be directly translatable to humans. Future studies could investigate whether a tissue-specific or a ubiquitous approach (which may be more representative of a patient’s molecular state) produces the same candidate modifiers, or whether scoring different phenotypes in SSIRDD models changes the same candidate modifiers. Modeling SSRIDDs in Drosophila was challenging due to the deleterious fitness effects associated with decreased expression of associated orthologs. The role of these genes in reproduction, healthspan, cancer, and other fitness traits is outside the scope of this dissertation, but these findings could indicate important disease modifiers or clues to disease pathogenesis. Assessment of cancer risk in SSRIDD Drosophila models would be an especially worthwhile future direction, as genes associated with SSRIDDs are also associated with cancers (Kadoch and Crabtree 2015). Clearly, the genotype-phenotype correlation for SSRIDD- and CdLS-associated genes is not well understood; much remains to be discovered in this field.

The studies included in this dissertation serve as a foundation for future work in higher order model organisms and provide starting points for studies using human data. Although findings from Drosophila models are not always translatable to humans, this dissertation uses Drosophila to answer questions that cannot be answered in human populations. Furthering the systems genetics approach taken here, collection and incorporation of additional layers of data, such as epigenome, proteome, and/or learning and memory behavioral data, will enrich our understanding of Drosophila models of neurodevelopmental disorders. Information gleaned from such animal models may lead to meaningful changes in disease diagnosis, management, and treatment for human patients.

# References

Bellen HJ, Wangler MF, Yamamoto S. The fruit fly at the interface of diagnosis and pathogenic mechanisms of rare and common human diseases. *Hum Mol Genet.* 2019. 28(R2):R207-14.

Bögershausen N, Wollnik B. Mutational landscapes and phenotypic spectrum of SWI/SNF-related intellectual disability disorders. *Front Mol Neurosci.* 2018. 11:252.

Bonilla M, McPherson M, Coreas J, Boulos M, Chavol P, Alrabadi RI, Loza-Coll M. Repeated ethanol intoxications of *Drosophila melanogaster* adults increases the resistance to ethanol of their progeny. *Alcohol Clin Exp Res.* 2021. 45(7):1370-82.

Branco J, Al-Ramahi I, Ukani L, Pérez AM, Fernandez-Funez P, Rincón-Limas D, Botas J. Comparative analysis of genetic modifiers in Drosophila points to common and distinct mechanisms of pathogenesis among polyglutamine diseases. *Hum Mol Genet.* 2008. 17(3):376-90.

Buff H, Smith AC, Korey CA. Genetic modifiers of Drosophila palmitoyl-protein thioesterase 1-induced degeneration. *Genetics.* 2007. 176(1):209-20.

Chow CY, Kelsey KJP, Wolfner MF, Clark AG. Candidate genetic modifiers of retinitis pigmentosa identified by exploiting natural variation in Drosophila. *Hum Mol Genet.* 2016. 25(4):651-9.

Chubak MC, Nixon KCJ, Stone MH, Raun N, Rice SL, Sarikahya M, Jones SG, Lyons TA, Jakub TE, Mainland RLM, et al. Individual components of the SWI/SNF chromatin remodelling complex have distinct roles in memory neurons of the Drosophila mushroom body. *Dis Model Mech.* 2019. 12(3):dmm037325.

Coll-Tane M, Krebbers A, Castells-Nobau A, Zweier C, Schenck A. Intellectual disability and autism spectrum disorders 'on the fly': Insights from Drosophila. *Dis Model Mech.* 2019. 12(5):dmm039180. doi: 10.1242/dmm.039180.

Cukier HN, Perez AM, Collins AL, Zhou Z, Zoghbi HY, Botas J. Genetic modifiers of *MeCP2* function in Drosophila. *PLoS Genet.* 2008. 4(9):e1000179.

Czekay DP, Kothe U. H/ACA small ribonucleoproteins: Structural and functional comparison between archaea and eukaryotes. *Front Microbiol.* 2021. 12:654370.

Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics.* 2011. 12(1):357.

Kadoch C, Crabtree GR. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv.* 2015. 1(5):e1500447.

Kawauchi S, Calof AL, Santos R, Lopez-Burks M, Young CM, Hoang MP, Chua A, Lao T, Lechner MS, Daniel JA, et al. Multiple organ system defects and transcriptional dysregulation in the *NIPBL* +/− mouse, a model of Cornelia de Lange syndrome. *PLOS Genet.* 2009. 5(9):e1000650.

Kline AD, Moss JF, Selicorni A, Bisgaard A, Deardorff MA, Gillett PM, Ishman SL, Kerr LM, Levin AV, Mulder PA, et al. Diagnosis and management of Cornelia de Lange syndrome: First international consensus statement. *Nat Rev Genet.* 2018. 19(10):649-66.

Lavoy S, Chittoor-Vinod V, Chow CY, Martin I. Genetic modifiers of neurodegeneration in a Drosophila model of Parkinson's disease. *Genetics.* 2018. 209(4):1345-56.

Liu J, Zhang Z, Bando M, Itoh T, Deardorff MA, Clark D, Kaur M, Tandy S, Kondoh T, Rappaport E, et al. Transcriptional dysregulation in *NIPBL* and cohesin mutant human cells. *PLoS Biol.* 2009. 7(5):e1000119.

Liu Y, Balaraman Y, Wang G, Nephew KP, Zhou FC. Alcohol exposure alters DNA methylation profiles in mouse embryos at early neurulation. *Epigenetics.* 2009. 4(7):500-11.

Mackay TFC, Huang W. Charting the genotype–phenotype map: Lessons from the *Drosophila melanogaster* genetic reference panel. *Wiley Interdiscip Rev Dev Biol.* 2018. 7(1):e289.

McClure KD, French RL, Heberlein U. A Drosophila model for fetal alcohol syndrome disorders: Role for the insulin pathway. *Dis Model Mech.* 2011. 4(3):335-46.

Mills JA, Herrera PS, Kaur M, Leo L, McEldrew D, Tintos-Hernandez JA, Rajagopalan R, Gagne A, Zhang Z, Ortiz-Gonzalez XR, et al. NIPBL +/- haploinsufficiency reveals a constellation of transcriptome disruptions in the pluripotent and cardiac states. *Sci Rep.* 2018. 8(1):1056-13.

Morozova TV, Huang W, Pray VA, Whitham T, Anholt RRH, Mackay TFC. Polymorphisms in early neurodevelopmental genes affect natural variation in alcohol sensitivity in adult Drosophila. *BMC Genomics.* 2015. 16(1):865.

Morozova TV, Hussain Y, McCoy LJ, Zhirnov EV, Davis MR, Pray VA, Lyman RA, Duncan LH, McMillen A, Jones A, et al. A *Cyclin E* centered genetic network contributes to alcohol-induced variation in Drosophila development. *G3 (Bethesda).* 2018. 8(8):2643-53.

Palu RAS, Ong E, Stevens K, Chung S, Owings KG, Goodman AG, Chow CY. Natural genetic variation screen in Drosophila identifies Wnt signaling, mitochondrial metabolism, and redox homeostasis genes as modifiers of apoptosis. *G3 (Bethesda).* 2019. 9(12):3995-4005.

Revenkova E, Focarelli ML, Susani L, Paulis M, Bassi MT, Mannini L, Frattini A, Delia D, Krantz I, Vezzoni P, et al. Cornelia de Lange syndrome mutations in *SMC1A* or *SMC3* affect binding to DNA. *Hum Mol Genet.* 2009. 18(3):418-27.

Singh CM, Heberlein U. Genetic control of acute ethanol-induced behaviors in drosophila. *Alcohol Clin Exp Res.* 2000. 24(8):1127-36.

Talsness DM, Owings KG, Coelho E, Mercenne G, Pleinis JM, Partha R, Hope KA, Zuberi AR, Clark NL, Lutz CM, et al. A Drosophila screen identifies *NKCC1* as a modifier of NGLY1 deficiency. *eLife.* 2020. 9.

Vasileiou G, Ekici A, Uebe S, Zweier C, Hoyer J, Engels H, Behrens J, Reis A, Hadjihannas M. Chromatin-remodeling-factor ARID1B represses wnt/β-catenin signaling. *Am J Hum Genet.* 2015. 97(3):445-56.

Wallén E, Auvinen P, Kaminen-Ahola N. The effects of early prenatal alcohol exposure on epigenome and embryonic development. *Genes.* 2021. 12(7):1095.

Weiss FD, Calderon L, Wang Y, Georgieva R, Guo Y, Cvetesic N, Kaur M, Dharmalingam G, Krantz ID, Lenhard B, et al. Neuronal genes deregulated in Cornelia de Lange syndrome respond to removal and re-expression of cohesin. *Nat Comm.* 2021. 12(1):2919.

Wu Y, Gause M, Xu D, Misulovin Z, Schaaf CA, Mosarla RC, Mannino E, Shannon M, Jones E, Shi M, et al. Drosophila *Nipped-B* mutants model Cornelia de Lange syndrome in growth and behavior. *PLoS Genet.* 2015. 11(11):e1005655.

Zhan L, Hanson KA, Kim SH, Tare A, Tibbetts RS. Identification of genetic modifiers of TDP-43 neurotoxicity in Drosophila. *PLoS ONE.* 2013. 8(2):e57214.

# APPENDICES

# Appendix A

## Supplementary material for Chapter 2

**All files are located at:**

https://clemson.box.com/s/ul80pg6zz382fzsukyxdodvepxi2ccy6

//MacPherson_dissertation/chapter02_ethanol_sedation_assay

*Supplementary Tables*

**Supplementary Table S2.1** Materials List

# Appendix B

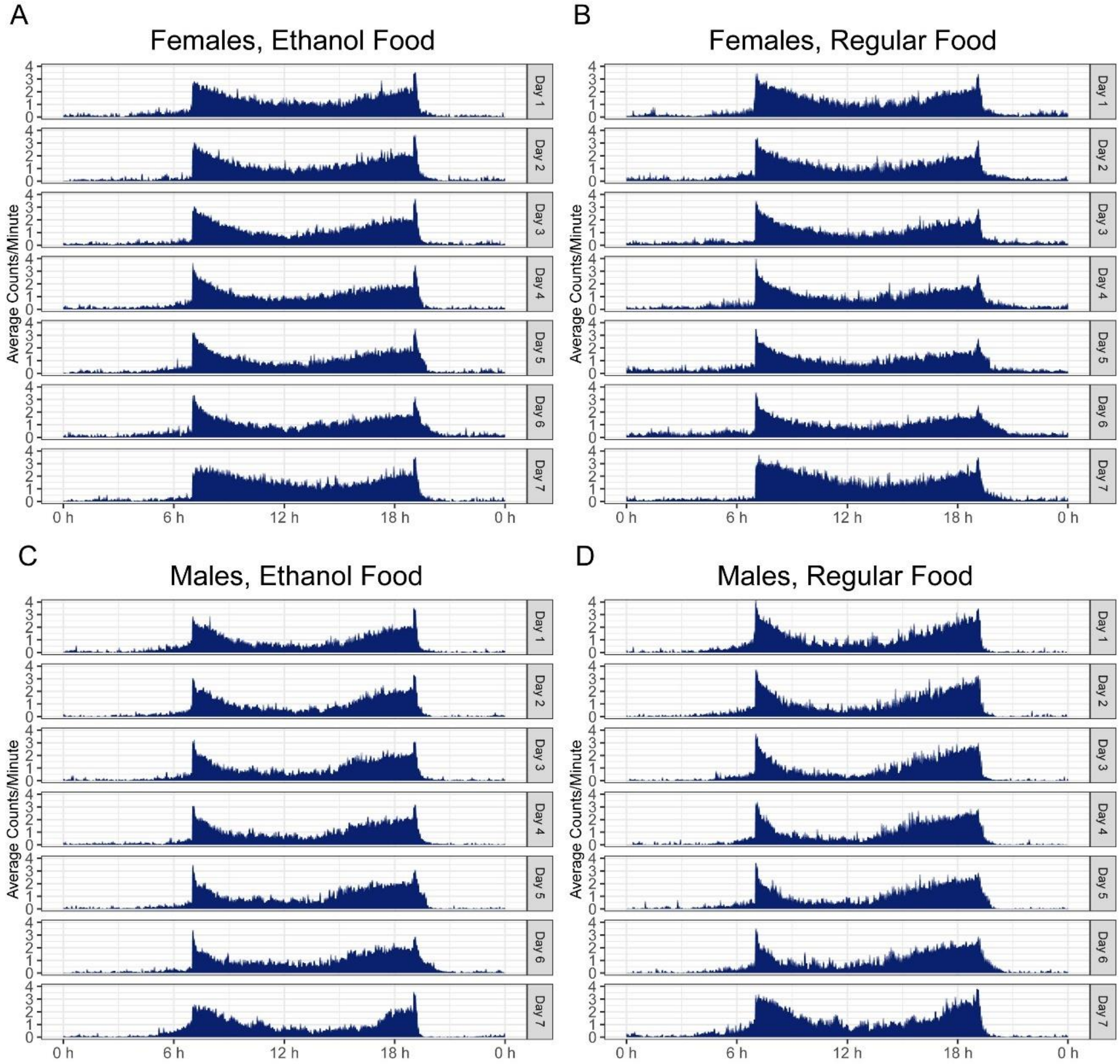
## Supplementary material for Chapter 3

**All files are located at:**

https://clemson.box.com/s/ul80pg6zz382fzsukyxdodvepxi2ccy6

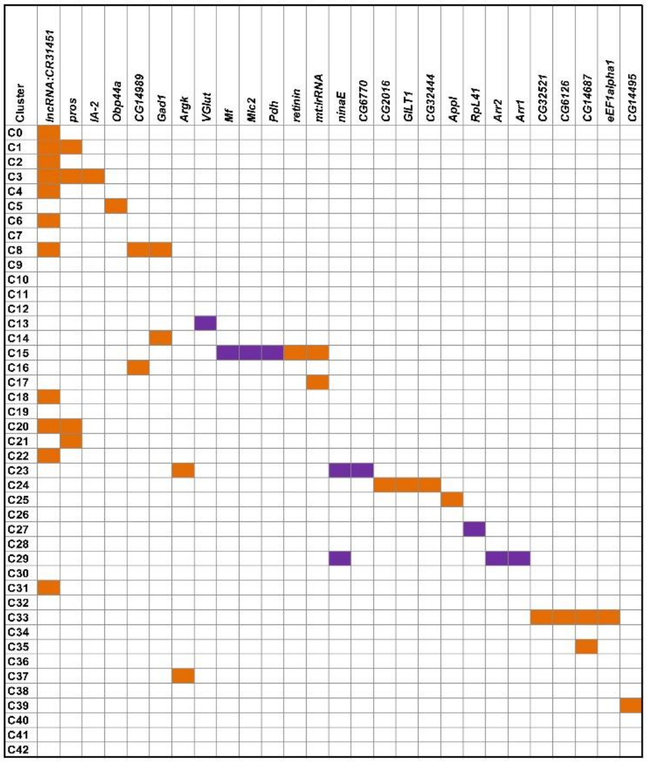
//MacPherson_dissertation/chapter03_scRNAseq

*Supplementary Figures*



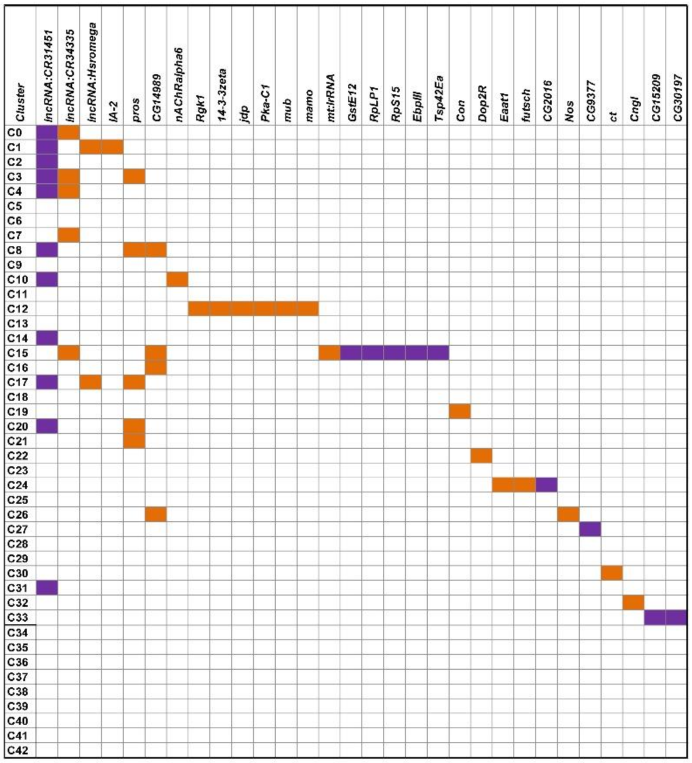
**Supplementary Figure S.3.1** Actograms showing average number of counts per fly per minute from females grown on (A) ethanol-supplemented food (10% v/v) and (B) regular food, and males grown on (C) ethanol-supplemented food (10% v/v) and (D) regular food. Actograms correspond to data collected for sleep and activity phenotypes shown in Figures 2D–G. Day hours are from 7:00 a.m. to 7:00 p.m., lights on 7 h after hour zero. Bin length = 5 min.

**A** Males



| Cluster | *lncRNA:CR31451* | *pros* | *IA-2* | *Obp44a* | *CG14989* | *Gad1* | *Argk* | *VGlut* | *Mf* | *Mlc2* | *Pdh* | *retinin* | *mt:lrRNA* | *ninaE* | *CG6770* | *CG2016* | *GILT1* | *CG32444* | *Appl* | *RpL41* | *Arr2* | *Arr1* | *CG32521* | *CG6126* | *CG14687* | *eEF1alpha1* | *CG14495* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C1 | 🟧 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | |
| C2 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C3 | 🟧 | 🟧 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | |
| C4 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C5 | | | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | |
| C6 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C7 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C8 | 🟧 | | | | 🟧 | 🟧 | | | | | | | | | | | | | | | | | | | | | |
| C9 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C10 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C11 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C12 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C13 | | | | | | | | 🟪 | | | | | | | | | | | | | | | | | | | |
| C14 | | | | | | 🟧 | | | | | | | | | | | | | | | | | | | | | |
| C15 | | | | | | | | | 🟪 | 🟪 | 🟪 | 🟧 | 🟧 | | | | | | | | | | | | | | |
| C16 | | | | | 🟧 | | | | | | | | | | | | | | | | | | | | | | |
| C17 | | | | | | | | | | | | | 🟧 | | | | | | | | | | | | | | |
| C18 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C19 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C20 | 🟧 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | |
| C21 | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | |
| C22 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C23 | | | | | | | 🟧 | | | | | | | 🟪 | 🟪 | | | | | | | | | | | | |
| C24 | | | | | | | | | | | | | | | | 🟧 | 🟧 | 🟧 | | | | | | | | | |
| C25 | | | | | | | | | | | | | | | | | | | 🟧 | | | | | | | | |
| C26 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C27 | | | | | | | | | | | | | | | | | | | | 🟪 | | | | | | | |
| C28 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C29 | | | | | | | | | | | | | | 🟪 | | | | | | | 🟪 | 🟪 | | | | | |
| C30 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C31 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C32 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C33 | | | | | | | | | | | | | | | | | | | | | | | 🟧 | 🟧 | 🟧 | 🟧 | |
| C34 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C35 | | | | | | | | | | | | | | | | | | | | | | | | | 🟧 | | |
| C36 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C37 | | | | | | | 🟧 | | | | | | | | | | | | | | | | | | | | |
| C38 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C39 | | | | | | | | | | | | | | | | | | | | | | | | | | | 🟧 |
| C40 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C41 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C42 | | | | | | | | | | | | | | | | | | | | | | | | | | | |

**B** Females



| Cluster | *lncRNA:CR31451* | *lncRNA:CR34335* | *lncRNA:Hsromega* | *IA-2* | *pros* | *CG14989* | *nAChRalpha6* | *Rgk1* | *14-3-3zeta* | *jdp* | *Pka-C1* | *mub* | *mamo* | *mt:lrRNA* | *GstE12* | *RpLP1* | *RpS15* | *EbpIII* | *Tsp42Ea* | *Con* | *Dop2R* | *Eaat1* | *futsch* | *CG2016* | *Nos* | *CG9377* | *ct* | *Cngl* | *CG15209* | *CG30197* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | 🟪 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C1 | 🟪 | | 🟧 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C2 | 🟪 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C3 | 🟪 | 🟧 | | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | |
| C4 | 🟪 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C7 | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C8 | 🟪 | | | | 🟧 | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | |
| C9 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C10 | 🟪 | | | | | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | |
| C11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C12 | | | | | | | | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | 🟧 | | | | | | | | | | | | | | | | | |
| C13 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C14 | 🟪 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C15 | | 🟧 | | | | 🟧 | | | | | | | | 🟧 | 🟪 | 🟪 | 🟪 | 🟪 | 🟪 | | | | | | | | | | | |
| C16 | | | | | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | |
| C17 | 🟪 | | 🟧 | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | |
| C18 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C19 | | | | | | | | | | | | | | | | | | | | 🟧 | | | | | | | | | | |
| C20 | 🟪 | | | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | |
| C21 | | | | | 🟧 | | | | | | | | | | | | | | | | | | | | | | | | | |
| C22 | | | | | | | | | | | | | | | | | | | | | 🟧 | | | | | | | | | |
| C23 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C24 | | | | | | | | | | | | | | | | | | | | | | 🟧 | 🟧 | 🟪 | | | | | | |
| C25 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C26 | | | | | | 🟧 | | | | | | | | | | | | | | | | | | | 🟧 | | | | | |
| C27 | | | | | | | | | | | | | | | | | | | | | | | | | | 🟪 | | | | |
| C28 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C29 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C30 | | | | | | | | | | | | | | | | | | | | | | | | | | | 🟧 | | | |
| C31 | 🟪 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C32 | | | | | | | | | | | | | | | | | | | | | | | | | | | | 🟧 | | |
| C33 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 🟪 | 🟪 |
| C34 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C35 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C36 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C37 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C38 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C39 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C40 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C41 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C42 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

**Supplementary Figure S.3.2** Differentially expressed genes across clusters in males (A) and females (B) after developmental alcohol exposure. Differentially expressed genes are listed on the top (columns) and cell clusters are represented by the rows. Upregulated genes are indicated with orange and downregulated genes are indicated with purple. Differentially expressed genes are filtered at $|\log_e FC| > 1.0$ and a Bonferroni adjusted $p < 0.05$.

*Supplementary Tables*

**Supplementary Table S3.1** ANOVA tables for viability, ethanol sensitivity, sleep, and activity.

**Supplementary Table S3.2** Sequencing statistics. F denotes females and M denotes males. C indicates control medium and E ethanol-supplemented medium. The numbers indicate replicates 1 and 2.

**Supplementary Table S3.3** Genes used to annotate cell clusters.

**Supplementary Table S3.4** List of differentially expressed genes in each cluster in males. Each sheet corresponds to the male analyses for the given cluster. "Avg_diff" is conditionally formatted to indicate up- and down-regulation of expression in ethanol compared to regular food (red: up-regulated, green: down-regulated and yellow: no difference). p_val: raw p-value from the differential expression analysis for the given gene in the corresponding cluster. avg_diff: the difference in the log(e) transformed average expression of the given gene in the corresponding cluster (sheet) between the two conditions (ethanol compared to regular food). Values above zero indicate up-regulation of expression due to developmental exposure to ethanol, and likewise, values below zero represent down-regulation of expression due to ethanol. p_val_adj: Bonferroni adjusted p-value. The DE matrix sheet is a summary of differentially expressed genes (columns) and the clusters in which they are differentially expressed (rows) with orange indicating upregulation and purple indicating downregulation at |avg_diff| thresholds of 0.25 and 1.

The All DE per cluster sheet and the All DE sheet are summaries of all the differentially expressed genes.

**Supplementary Table S3.5** List of differentially expressed genes in each cluster in females. Each sheet corresponds to the female analyses for the given cluster. “Avg_diff” is conditionally formatted to indicate up- and down-regulation of expression in ethanol compared to regular food (red: up-regulated, green: down-regulated and yellow: no difference). p_val: raw p-value from the differential expression analysis for the given gene in the corresponding cluster. avg_diff: the difference in the log(e) transformed average expression of the given gene in the corresponding cluster (sheet) between the two conditions (ethanol compared to regular food). Values above zero indicate up-regulation of expression due to developmental exposure to ethanol, and likewise, values below zero represent down-regulation of expression due to ethanol. p_val_adj: Bonferroni adjusted p-value. The DE matrix sheet is a summary of differentially expressed genes (columns) and the clusters in which they are differentially expressed (rows) with orange indicating upregulation and purple indicating downregulation at |avg_diff| thresholds of 0.25 and 1. The All DE per cluster sheet and the All DE sheet are summaries of all the differentially expressed genes.

**Supplementary Table S3.6** Human orthologs of differentially expressed genes.

**Supplementary Table S3.7** Common differentially expressed genes upon developmental alcohol exposure and acute exposure to cocaine.

**Supplementary Table S3.8** Comparison of cell type-specific differentially expressed genes between developmental ethanol exposure and acute cocaine exposure. Meta-comparison sheet contains the mapping of clusters and cell types between the two datasets as well as the methodology and summary of the comparisons. The rest of the sheets contain the list of statistically significantly differentially expressed genes, their Loge fold change values, the calculations of the comparisons between the two datasets for each cell type-category. The comparisons were done for each cell type-category separately for the male and female datasets.

**Supplementary Table S3.9** Gene ontology analysis of differentially expressed genes identified both after developmental exposure to alcohol and acute intake of cocaine.

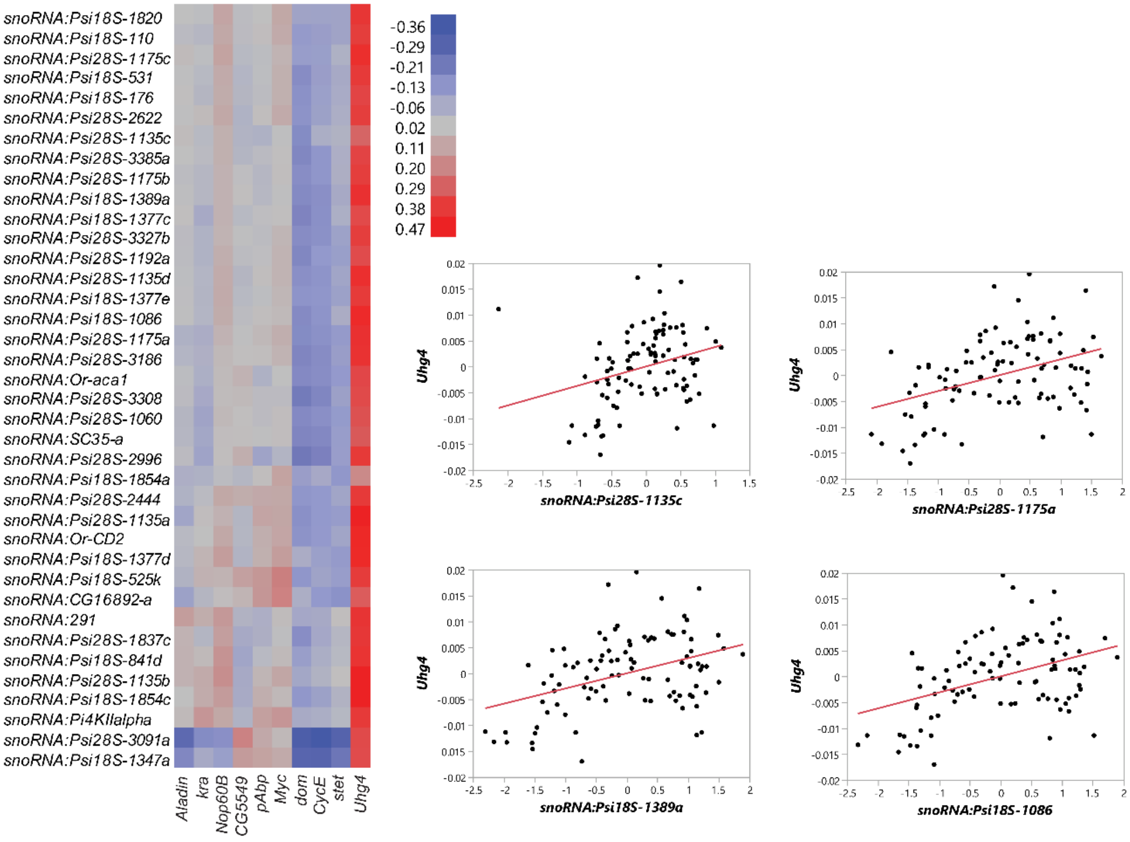
# Appendix C

## Supplementary material for Chapter 4

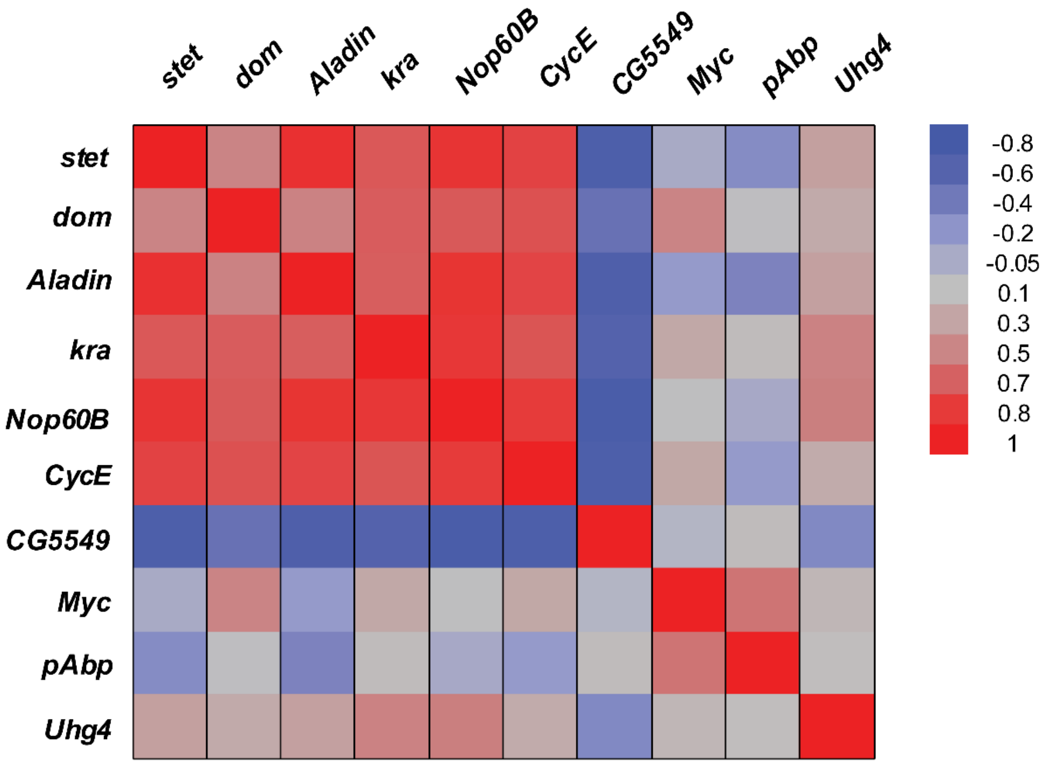
**All files are located at:**

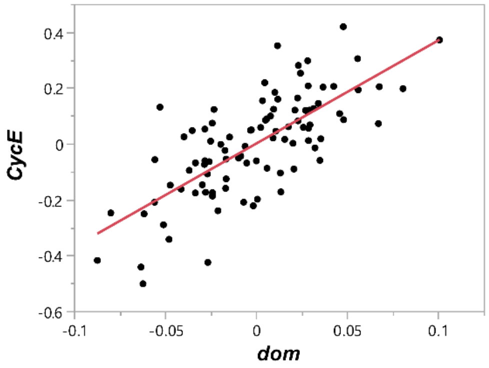
https://clemson.box.com/s/ul80pg6zz382fzsukyxdodvepxi2ccy6

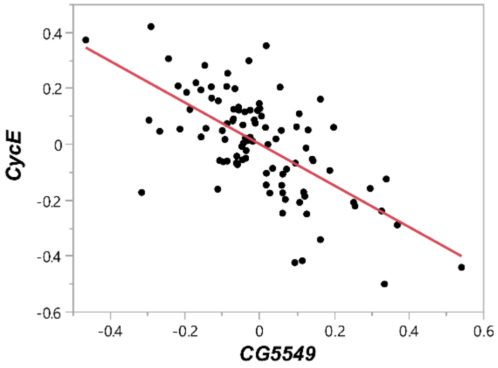
//MacPherson_dissertation/chapter04_FASD_DGRP

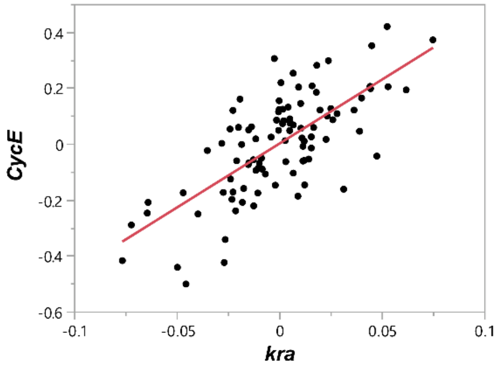


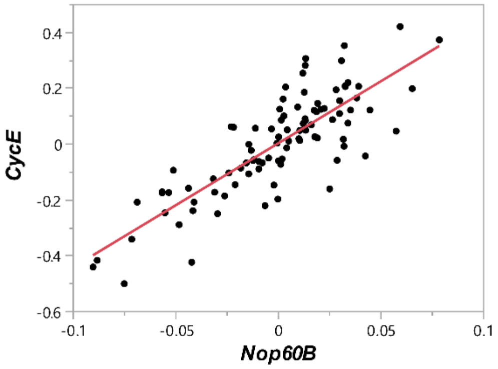
**Supplementary Figure S4.1.** Correlations between variation in expression of snoRNAs after chronic exposure to ethanol and variation in expression of their host genes. *CycE* is included since it is a hub gene in a genetic network associated with variation in ethanol-induced variation in development time and viability [16], and *Myc* is included as it is associated with ribosome biogenesis and has been implicated as a regulator of *Uhg4* [38]. The graphs on the right illustrate examples of scatter plots of the correlations between expression of *Uhg4* and several snoRNAs.


stet
dom
Aladin
kra
Nop60B
CycE
CG5549
Myc
pAbp
Uhg4
-0.8
-0.6
-0.4
-0.2
-0.05
0.1
0.3
0.5
0.7
0.8
1


CycE
dom
0.4
0.2
0
-0.2
-0.4
-0.6
-0.1
-0.05
0
0.05
0.1


CycE
CG5549
0.4
0.2
0
-0.2
-0.4
-0.6
-0.4
-0.2
0
0.2
0.4
0.6


CycE
kra
0.4
0.2
0
-0.2
-0.4
-0.6
-0.1
-0.05
0
0.05
0.1


CycE
Nop60B
0.4
0.2
0
-0.2
-0.4
-0.6
-0.1
-0.05
0
0.05
0.1

**Supplementary Figure S4.2.** Correlations between ethanol-induced variation in expression of snoRNA host genes and variation in expression of *CycE*. The scatter plots illustrate examples of correlations between expression of *CycE* and several snoRNA host genes.

*Supplementary Files*

**Supplementary File S4.1.** DGRP lines used for transcriptional profiling

**Supplementary File S4.2.** Three way mixed model ANOVA for differences in transcript abundances between flies reared on regular and ethanol supplemented medium for sexes combined. The main effects are treatment (fixed, reared on standard vs. ethanol supplemented medium), Sex (fixed, males and females) and DGRP Line (random). All interactions with Line are random, the others are fixed.

**Supplementary File S4.3.** ANOVA for differences in transcript abundances between flies reared on regular and ethanol supplemented medium for (A) females and (B) males.

**Supplementary File S4.4.** Genes with significant *LxT* interactions in transcript abundances. (A) Results of female-only mixed-effect models for all expressed gene profiles, including alignment bias estimates. (B) Results of male-only mixed-effect models for all expressed gene profiles, including alignment bias estimates. (C) Genes with significant *LxT* interactions common to females and males and that are female-specific or male-specific. (D) Enrichment analyses for genes with significant *LxT* interactions in transcript abundances in females. (E) Enrichment analyses for genes with significant *LxT* interactions in transcript abundances in males. (F) Enrichment analyses for genes with significant *LxT* interactions in transcript abundances in both sexes. (G) Enrichment analyses for genes with significant *LxT* interactions in transcript abundances in females only. (H) Enrichment analyses for genes with significant *LxT* interactions in transcript abundances in males only. (I) Human orthologs for genes with significant *LxT* interactions in transcript abundances.

**Supplementary File S4.5.** (A) eQTL mapping and model selection results for genes with a statistically significant *L*x*T* ANOVA term in females and males. (B) Genes containing eQTLs (in the gene body or within a 1,000 bp window of the eQTL) in females and males. Intergenic: > 1,000 bp from any gene body.

**Supplementary File S4.6.** snoRNAs with altered gene expression following chronic exposure to ethanol during development and their host genes. snoRNAs that occur in clusters without intervening genes are in bold font and square brackets.

**Supplementary File S4.7** Mixed model ANOVAs for sleep and activity phenotypes. Line, Sex, and Treatment are fixed effects, Replicate (Rep) is random. "df": degrees of freedom; "SS": Type III Sums of Squares; "MS": Mean square.

# Appendix D

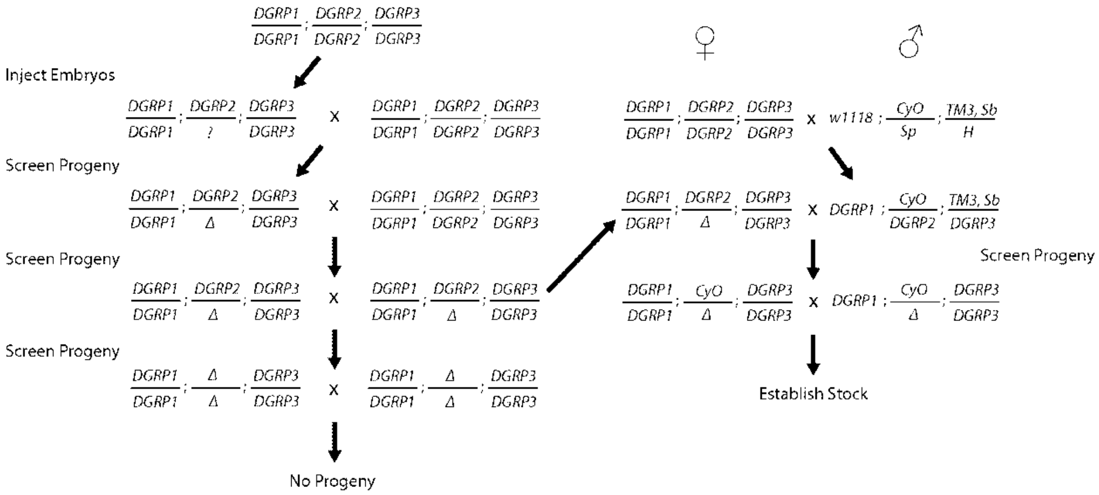
## Supplementary material for Chapter 5

**All files are located at :**

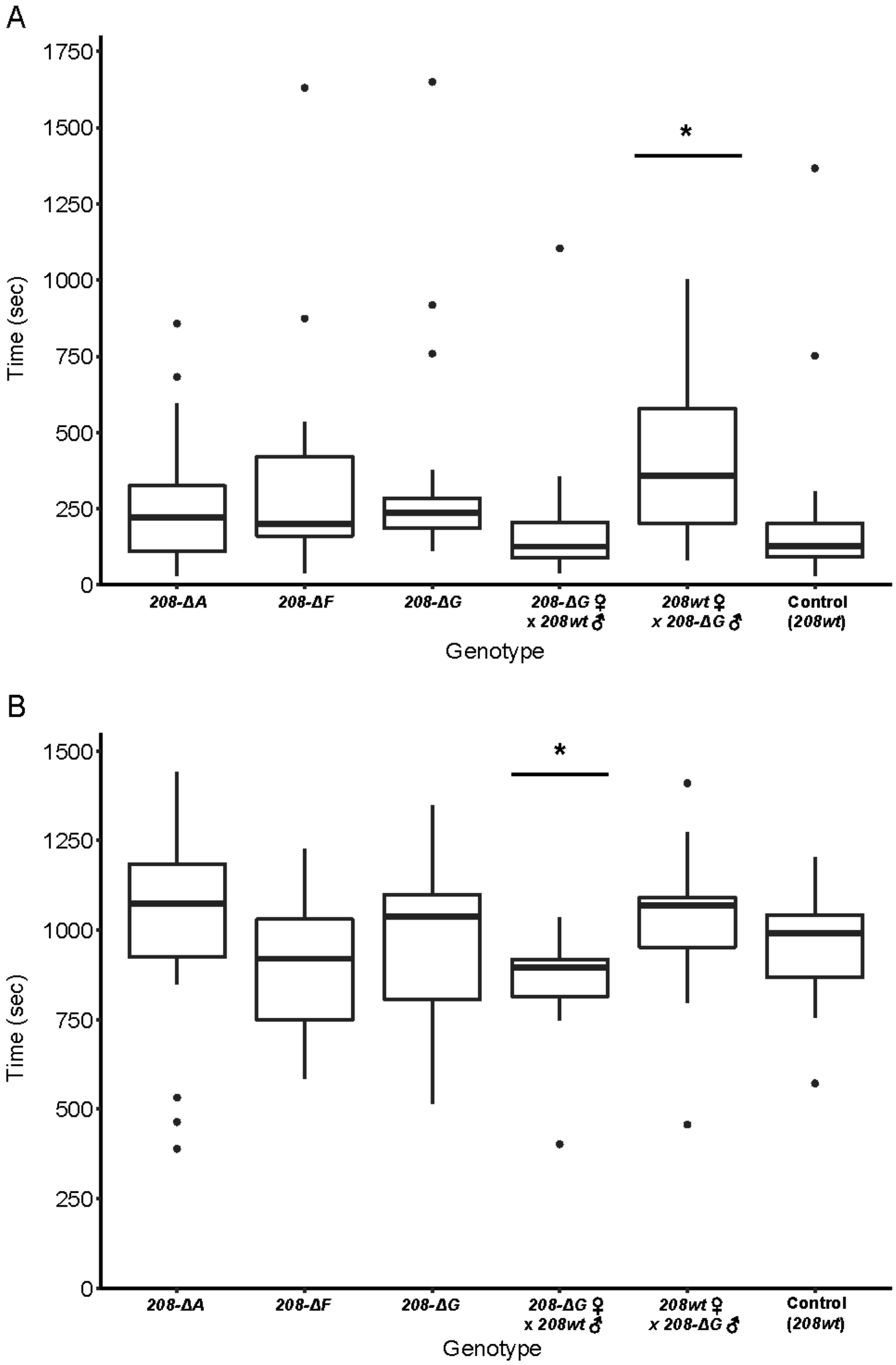
https://clemson.box.com/s/ul80pg6zz382fzsukyxdodvepxi2ccy6

//MacPherson_dissertation/chapter05_uhg4

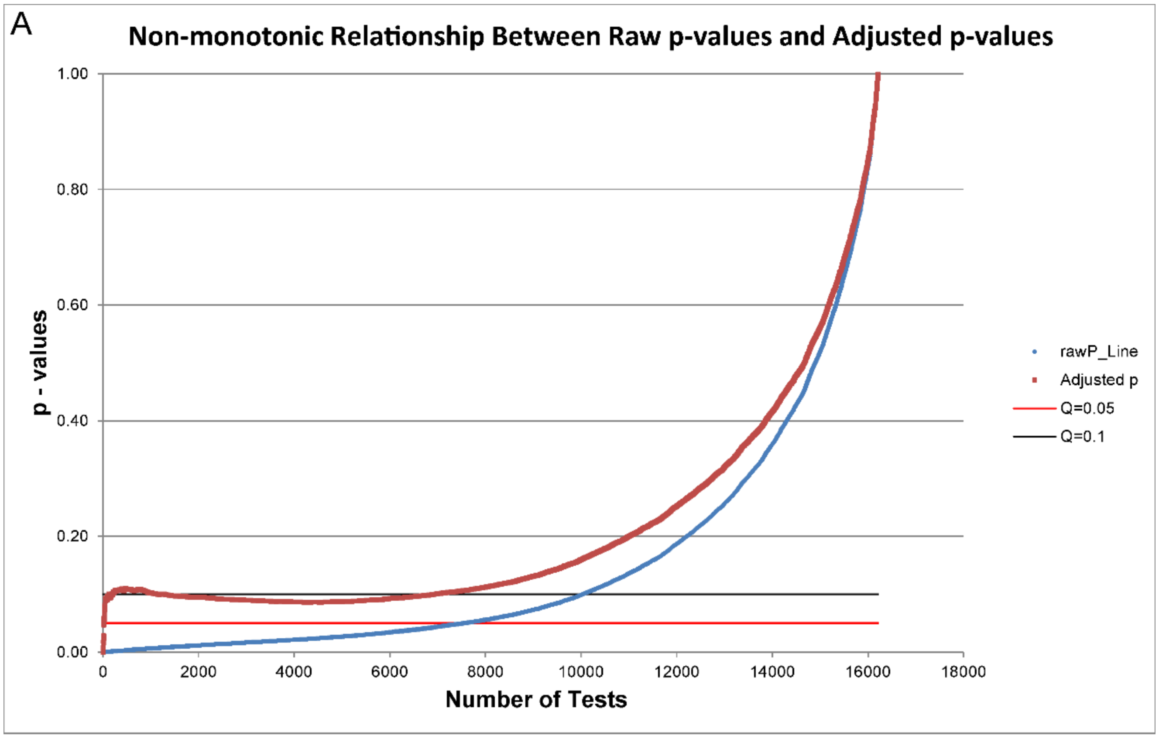
*Supplementary Figures*

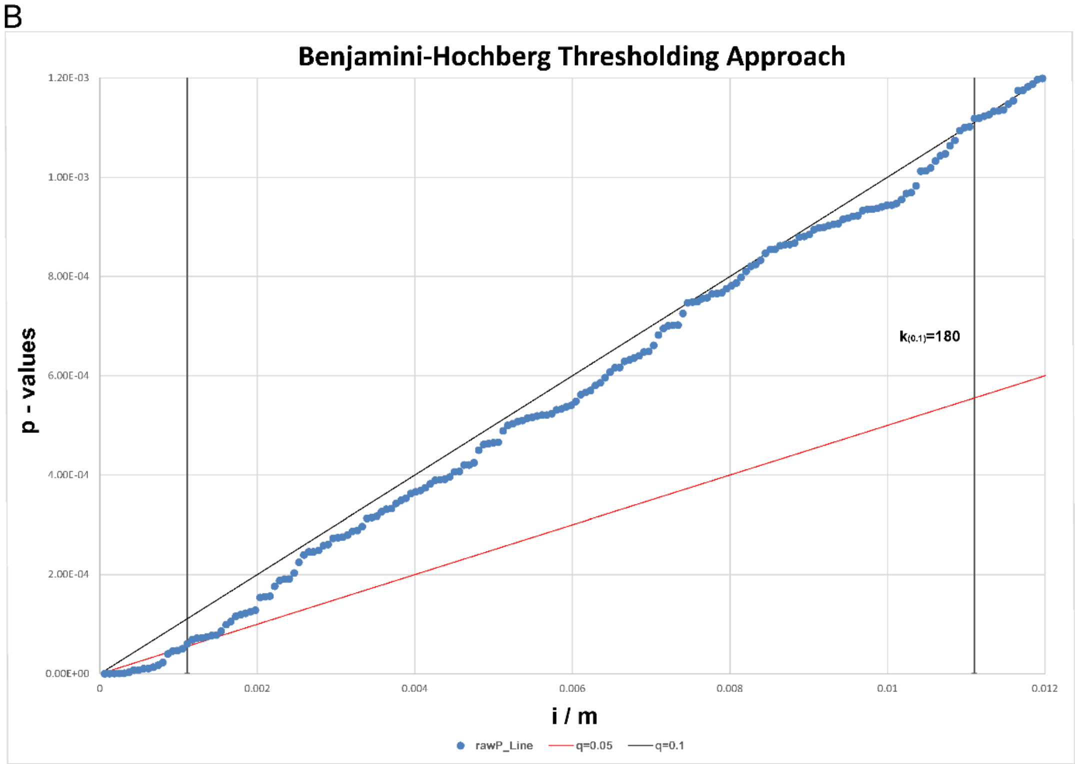

♀
♂
Inject Embryos
Screen Progeny
Screen Progeny
Screen Progeny
DGRP1/DGRP1 ; DGRP2/DGRP2 ; DGRP3/DGRP3
DGRP1/DGRP1 ; DGRP2/? ; DGRP3/DGRP3 x DGRP1/DGRP1 ; DGRP2/DGRP2 ; DGRP3/DGRP3
DGRP1/DGRP1 ; DGRP2/Δ ; DGRP3/DGRP3 x DGRP1/DGRP1 ; DGRP2/DGRP2 ; DGRP3/DGRP3
DGRP1/DGRP1 ; DGRP2/Δ ; DGRP3/DGRP3 x DGRP1/DGRP1 ; DGRP2/Δ ; DGRP3/DGRP3
DGRP1/DGRP1 ; Δ/Δ ; DGRP3/DGRP3 x DGRP1/DGRP1 ; Δ/Δ ; DGRP3/DGRP3
No Progeny
DGRP1/DGRP1 ; DGRP2/DGRP2 ; DGRP3/DGRP3 x w1118 ; CyO/Sp ; TM3, Sb/H
DGRP1/DGRP1 ; DGRP2/Δ ; DGRP3/DGRP3 x DGRP1 ; CyO/DGRP2 ; TM3, Sb/DGRP3
Screen Progeny
DGRP1/DGRP1 ; CyO/Δ ; DGRP3/DGRP3 x DGRP1 ; CyO/Δ ; DGRP3/DGRP3
Establish Stock

**Supplementary Figure S5.1.** Crossing scheme to generate *Uhg4* deletion CRISPR mutants. Following injections of *Cas9* and gRNA vectors into embryos of each DGRP line, resulting progeny were screened for presence of a deletion around *Uhg4*, and if a deletion was present, crossed to the original genetic background to generate additional flies heterozygous for the original deletion. Backcrossing siblings heterozygous for the same mutation resulted in homozygous flies that were sterile for all isolated deletion mutations. Virgin female flies heterozygous for a specific deletion were then crossed to male flies containing the *CyO* balancer. Resulting progeny were screened for the presence of the *CyO* balancer chromosome as well as the respective mutation and crossed to full siblings to establish the stock. All females were crossed as virgin flies. “*Δ*” refers to a specific *Uhg4* deletion.

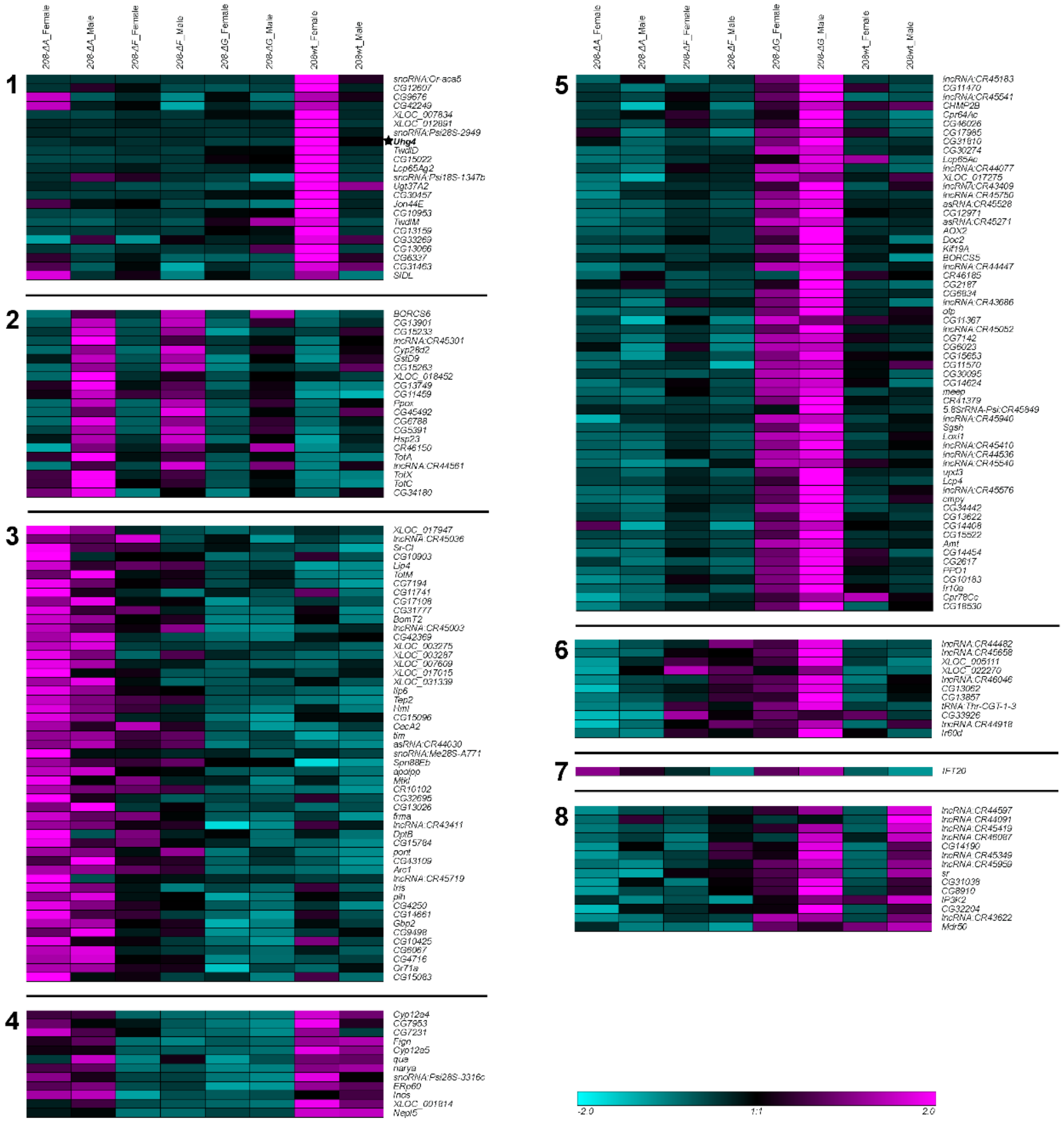


**Supplementary Figure S5.2.** Mating phenotypes of *Uhg4* deletion flies. Boxplots displaying time, in seconds, of (A) mating latency (time until mating begins) and (B) mating duration (length of copulation) for each of the *Uhg4* mutant lines (*208-ΔA*, *208-ΔF*, *208-ΔG*), the control line DGRP_208 (*208wt*), and a representative mutant (*208-ΔG*) versus DGRP_208 wildtype pairings (*208-ΔG* females x *208wt* males, *208wt* females x *208-ΔG* males). N = 22-24 pairings of 3-5 day old virgin flies per line. Only flies which successfully initiated or completed mating within 30 minutes were included in analysis for mating latency and mating duration, respectively. See Supplementary Table S5.1A. * $p<0.05$

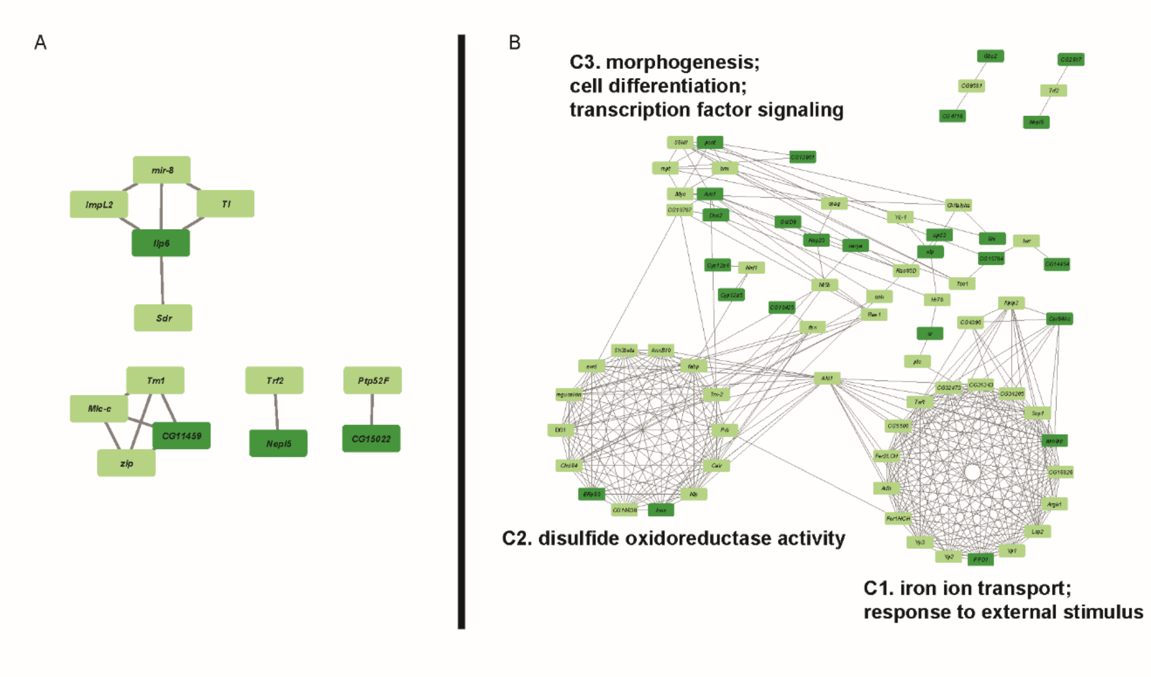

A
Non-monotonic Relationship Between Raw p-values and Adjusted p-values
p - values
1.00
0.80
0.60
0.40
0.20
0.00
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
Number of Tests
rawP_Line
Adjusted p
Q=0.05
Q=0.1


B
Benjamini-Hochberg Thresholding Approach
p - values
1.20E-03
1.00E-03
8.00E-04
6.00E-04
4.00E-04
2.00E-04
0.00E+00
0
0.002
0.004
0.006
0.008
0.01
0.012
i / m
k(0.1)=180
rawP_Line
q=0.05
q=0.1

**Supplementary Figure S5.3.** Non-monotonic relationship between raw and Benjamini-Hochberg adjusted p–values. (A) raw *p*-values plotted against number of tests; (B) BH-FDR thresholds on raw *p*-values.  Raw and adjusted *p*-values are shown in blue and red, respectively. “i”: rank, “m”: number of tests. See corresponding Supplementary Table S5.2.



**Supplementary Figure S5.4** K-means clusters. K-means clusters derived from genes and NTRs with differential expression in a global transcriptomic analysis using genes significant for the *Line* or *Line*×*Sex* terms (BH-FDR < 0.1). *Uhg4* is shown in bold and indicated with a star symbol. Magenta indicates a relatively higher degree of expression, blue indicates a relatively lower degree of expression.


A
mir-8
ImpL2
Tl
Ilp6
Sdr
Tm1
Mlc-c
CG11459
zip
Trf2
Nepl5
Ptp52F
CG15022
B
C3. morphogenesis;
cell differentiation;
transcription factor signaling
Gbp2
CG2817
CG9581
Trf2
CG4716
Nepl5
pont
CG13901
Myo
GstD9
Hsp23
Cyp12a4
Cyp12a5
CG10425
Rac85D
Hr78
CG15784
CG14454
CG4390
Sh3beta
AnxB10
fabp
Ald1
regucalcin
Trx-2
Pdi
Calr
ERp60
CG32473
CG31343
CG34205
Tsf1
Scp1
CG5550
Fer2LCH
Fer1HCH
CG16826
Lsp2
Yp1
Yp2
Yp3
PPO1
C2. disulfide oxidoreductase activity
C1. iron ion transport;
response to external stimulus

**Supplementary Figure S5.5.** Additional interaction networks. Interaction networks containing physical and genetic interactions with genes with a significant *Line* effect. (A) Networks generated from an input of 20 differentially expressed genes/NTRs (BH-FDR < 0.05) including neighbors within at least 1 degree. (B) Networks generated from an input of 180 coregulated genes (BH-FDR < 0.1) including neighbors within at least 2 degrees. Annotation is based on enriched Gene Ontology terms. Dark green indicates the genes in the input data set and light green indicates interaction neighbors. Names are Drosophila gene symbols. See Supplementary Table S5.4.

*Supplementary Tables*

**Supplementary Table S5.1.** Analyses of behavioral phenotypes. Analyses of variance (ANOVA) and Fisher's exact tests comparing *Uhg4* deletion flies to their respective controls.

**Supplementary Table S5.2.** Differentially expressed genes. (A) BH-FDR correction for genes based on *Line* and (B) *LinexSex* model terms*;* (C) Normalized average read counts values for the 180 coregulated genes and 13 NTRs; (D) FDR correction for NTRs genes based on *Line* and *LinexSex* model terms. See corresponding Supplementary Figure S5.3.

**Supplementary Table S5.3.** Analysis of differential expression. For each gene (A) and NTR (B), the FBgn number or NTR ID, raw *p*-values, BH-FDR adjusted *p*-values for the effects of *Line* (*L*), *Sex* (*S*), and *Line*×*Sex* (*L*×*S*) according to the ANOVA model $Y = \mu + L + S + L \times S + \varepsilon$.

**Supplementary Table S5.4.** Enriched Gene Ontology (GO) Biological Process, Molecular Function and REACTOME terms and their raw and BH-FDR adjusted *p*-values. Groups of genes analyzed include 20 genes/NTRs with differential expression for the *Line* or *Line*×*Sex* terms (BH-FDR < 0.05), 20 genes/NTRs and their 1-degree neighbors, 193 coregulated genes/NTRs (BH-FDR < 0.1) due to a significant *Line* or *Line*×*Sex* effect, 193 genes/NTRs and their 1-degree neighbors, genes within each

MCODE cluster (Figure 5.4, Supplementary Figure S5.5), and genes in each k-means cluster (Supplementary Figure S5.4).

**Supplementary Table S5.5.** Gene lists. Lists of gene names, symbols, and Flybase IDs (FBID) for each network cluster and k-means cluster.

*Supplementary Files*

**Supplementary File S5.1.** Video of a representative *Uhg4* deletion (*208-ΔG*) female fly displaying an elevated, splayed resting wing position compared to a wildtype female fly. The *Uhg4* deletion fly is walking during the first part of the video and the wildtype fly is walking at the end of the video.

**Supplementary File S5.2.** Video of a representative *Uhg4* deletion (*208-ΔG*) male fly displaying an elevated, splayed resting wing position compared to a wildtype male fly. The *Uhg4* deletion fly is to the right of the wildtype fly throughout the video.

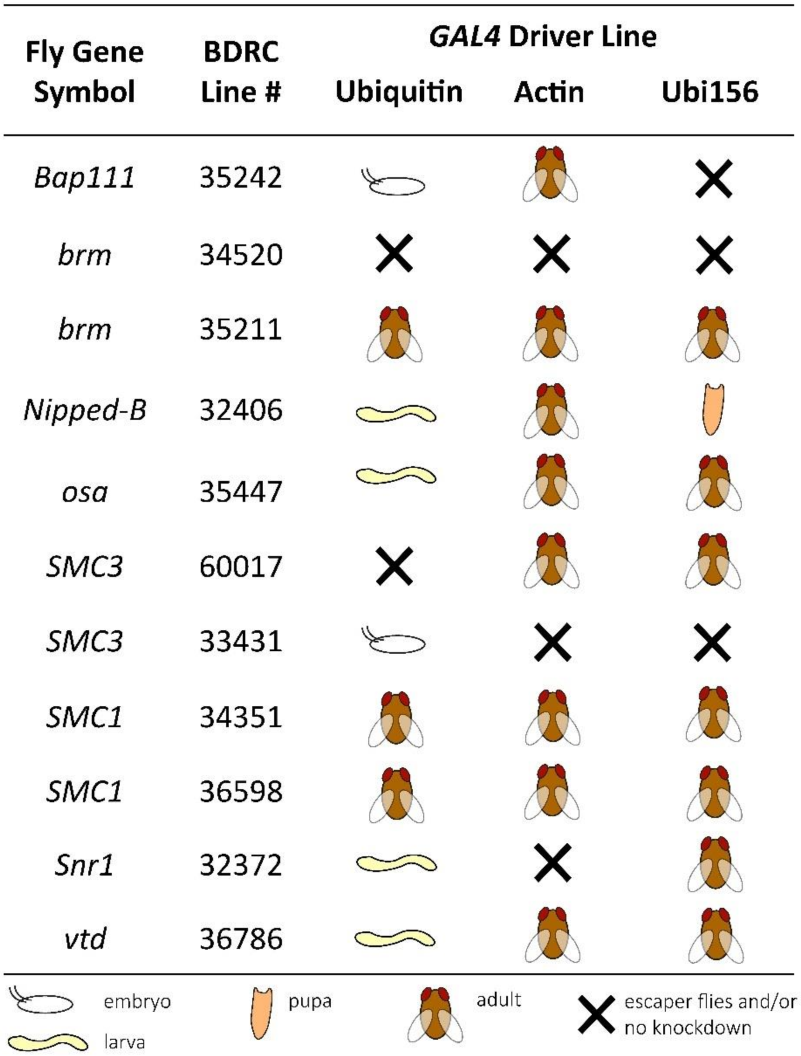
# Appendix E

## Supplementary material for Chapter 6

**All files are located at :**

https://clemson.box.com/s/ul80pg6zz382fzsukyxdodvepxi2ccy6

//MacPherson_dissertation/chapter06_SSRIDDs_CdLS

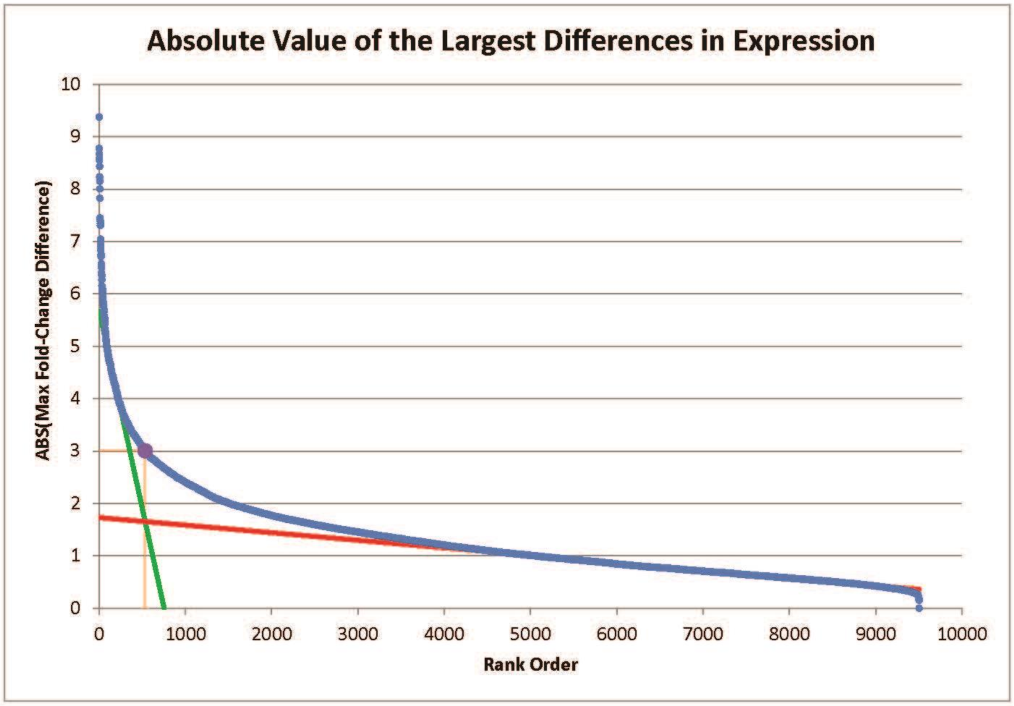


| Fly Gene Symbol | BDRC Line # | *GAL4* Driver Line | | |
|---|---|---|---|---|
| | | Ubiquitin | Actin | Ubi156 |
| *Bap111* | 35242 | 🥚 | 🪰 | ✕ |
| *brm* | 34520 | ✕ | ✕ | ✕ |
| *brm* | 35211 | 🪰 | 🪰 | 🪰 |
| *Nipped-B* | 32406 | 🐛 | 🪰 | 🫘 |
| *osa* | 35447 | 🐛 | 🪰 | 🪰 |
| *SMC3* | 60017 | ✕ | 🪰 | 🪰 |
| *SMC3* | 33431 | 🥚 | ✕ | ✕ |
| *SMC1* | 34351 | 🪰 | 🪰 | 🪰 |
| *SMC1* | 36598 | 🪰 | 🪰 | 🪰 |
| *Snr1* | 32372 | 🐛 | ✕ | 🪰 |
| *vtd* | 36786 | 🐛 | 🪰 | 🪰 |

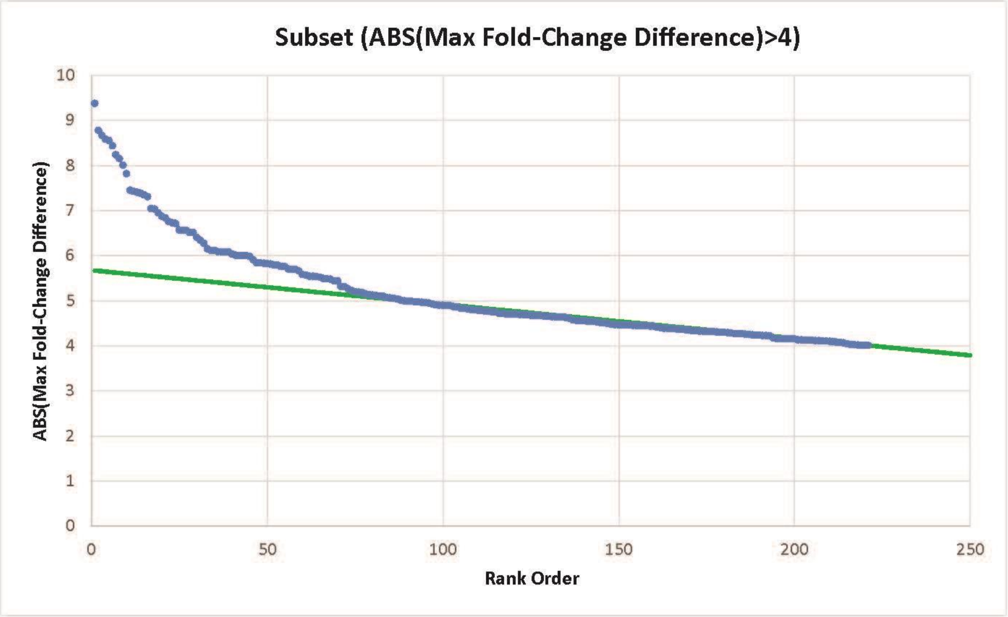
🥚 embryo
🐛 larva
🫘 pupa
🪰 adult
✕ escaper flies and/or no knockdown

**Supplementary Figure S6.1. Gross viability observations in potential CSS/NCBRS and CdLS fly models.** Life stage shown is the final stage of the *Drosophila* life cycle where live individuals were observed. "X" indicates flies did not have detectable levels of gene knockdown, as quantified via qRT-PCR.

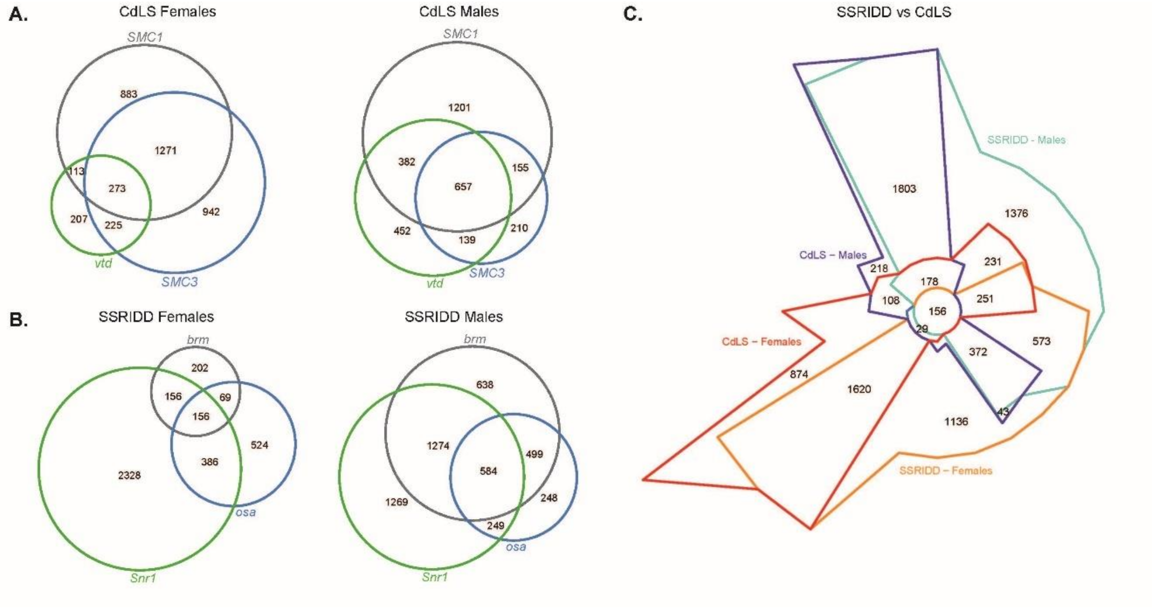
A


Absolute Value of the Largest Differences in Expression
ABS(Max Fold-Change Difference)
10
9
8
7
6
5
4
3
2
1
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Rank Order

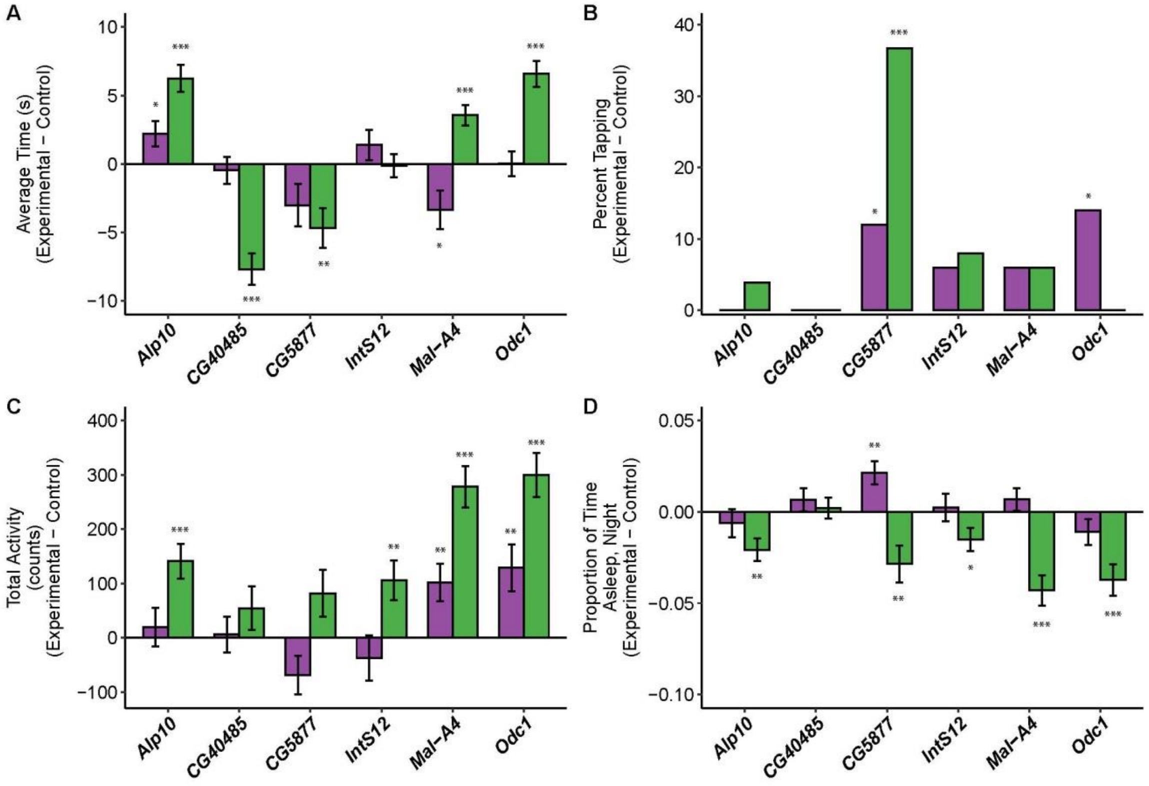
B


Subset (ABS(Max Fold-Change Difference)>4)
ABS(Max Fold-Change Difference)
10
9
8
7
6
5
4
3
2
1
0
0
50
100
150
200
250
Rank Order

**Supplementary Figure S6.2. Selection of genes for k-means clustering.** Elbow plots of maximal fold change in expression plotted against rank order (blue) across all analyses for each of 9657 genes (A) and for genes with a maximum fold change difference greater than 4 (B). See Table S7. The red and green lines were fit to roughly linear segments of the generated distribution (blue). The orange lines are drawn from the plot elbow (determined by the x coordinate of the intersection of the green and red lines) to the x and y axes.


A.
CdLS Females
SMC1
883
1271
113
273
207
225
942
vtd
SMC3
CdLS Males
SMC1
1201
382
657
155
452
139
210
vtd
SMC3
B.
SSRIDD Females
brm
202
156
69
156
2328
386
524
osa
Snr1
SSRIDD Males
brm
638
1274
584
499
1269
249
248
osa
Snr1
C.
SSRIDD vs CdLS
SSRIDD - Males
1803
1376
CdLS – Males
218
178
231
108
156
251
29
CdLS – Females
372
573
874
1620
43
1136
SSRIDD – Females

**Supplementary Figure S6.3. Overlap of differentially expressed genes in SSRIDD and CdLS fly models.** Venn diagrams displaying the number of differentially expressed genes (FDR < 0.05), in **SSRIDD** and CdLS fly models, sexes separately. Pairwise gene-specific analyses from (A) CdLS fly models and (B) **SSRIDD** fly models. Panel (C) shows overlap of disease-specific analyses, pooled across disease-associated genes.



**Supplementary Figure S6.4. Altered phenotypes due to knockdown of co-regulated genes.** Bar plots displaying differences in the average values of the experimental line versus the control line for (A) startle response, (B) percent of flies tapping, (C) total activity, and (D) proportion of time asleep at night. All lines have *Ubi156-GAL4*-mediated RNAi knockdown. Females and males are shown in purple and green, respectively. See Table S14 for ANOVAs (A,B,D) and Fishers Exact Tests (C). N=29-32 per sex per line. Error bars represent standard error of the difference based on error propagation (Burns and Dobson 1981). Asterisks represent pairwise analyses of the experimental line vs the control, sexes separately. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

*Supplementary Tables*

**Supplementary Table S6.1. Fly reagents and primer sequences.** Drosophila reagents and primer sequences. (A) Drosophila lines used. (B) Primer sequences used for qRT-PCR. BDSC: Bloomington Drosophila Stock Center.

**Supplementary Table S6.2. Ortholog prediction scores for potential focal genes.** Human-Drosophila ortholog prediction scores generated using Drosophila RNAi Screening Center Integrative Ortholog Prediction Tool (DIOPT). Human genes associated with SSRIDDs and Cornelia de Lange syndrome.

**Supplementary Table S6.3. Percent knockdown of focal genes.** Average RNAi-mediated qRT-PCR knockdown of focal genes.

**Supplementary Table S6.4. Quantification of changes in behavior and brain morphology from knockdown of focal genes.** Quantification of changes in behavior and brain morphology from RNAi knockdown. Statistical analyses characterizing SSRIDD and CdLS fly models. (A) ANOVAs for startle response. (B) Fisher's Exact Tests for tapping behavior. (C) ANOVAs for sleep and activity measurements. (D) ANOVAS for mushroom body lobe lengths. (E) Levene's and Brown-Forsythe Tests for unequal variances of mushroom body lobe length data. (F) Gross brain abnormalities. Line and Sex are fixed effects. df: degrees of freedom, SS: Type III Sum of Squares, MS: Mean Squares.

**Supplementary Table S6.5. ANOVA results from differential expression analyses.** Gene name, gene symbol, FlyBase ID, normalized read counts (counts per million), and raw and Benjamini-Hochberg FDR adjusted *p*-values for all genes for all model terms used in the ANOVA analyses. (A) Full model using all knockdown lines and the control according to the model $Y = \mu + Line + Sex + Line \times Sex + \varepsilon$ for 15915 genes. (B-G) Pairwise comparisons of single gene knockdown vs. the control (sexes together $Y = \mu + Line + Sex + Line \times Sex + \varepsilon$; and sexes separately $Y = \mu + Line + \varepsilon$) on the 9657 genes from the full model differentially expressed (FDR < 0.05) for the *Line* and/or *Line x Sex* terms. (B) *brm*. (C) *osa*. (D) *Snr1*. (E) *SMC1*. (F) *SMC3*. (G) *vtd*. (H-I) Disease-specific comparisons (sexes together $Y = \mu + Line + Sex + Line \times Sex + \varepsilon$; and sexes separately $Y = \mu + Line + \varepsilon$). (H) SSRIDDs. (I) Cornelia de Lange syndrome (CdLS).

**Supplementary Table S6.6. Overlap of differentially expressed genes across analyses.** FDR-corrected *p*-values less than 0.05 for the *Line* term of each of the 9657 genes. (A) Pairwise analyses of each knockdown line compared to the control, sexes separately. (B) Disease-specific analyses, sexes separately. NA indicates FDR-corrected *P*-values for the effect of *Line* greater than 0.05.

**Supplementary Table S6.7. k-means threshold.** (A) Average log2 fold change values for each differentially expressed gene for each set of samples, as well as maximum, minimum across all samples. (B) Determination of threshold by ranking, indexing and fitting lines to fold change plots. fc: log2 fold change; f: females, m: males.

**Supplementary Table S6.8. k-means clustering gene lists.** Lists of genes within each k-means cluster. (A) Females. (B) Males.

**Supplementary Table S6.9. Gene Ontology (GO) analyses for differentially expressed genes.** “Analysis” indicates the gene set used in the analysis.

**Supplementary Table S6.10. Gene Ontology (GO) analyses for k-means clusters.** “Analysis” indicates the gene set used in the analysis.

**Supplementary Table S6.11. Ortholog prediction scores for differentially expressed genes**. Drosophila-human ortholog prediction scores, generated using Drosophila RNAi Screening Center Integrative Ortholog Prediction Tool (DIOPT). Differentially expressed fly genes for each by-sex pairwise comparison.

**Supplementary Table S6.12. Ortholog prediction scores and known disease associations for co-regulated genes.** Drosophila-human ortholog prediction scores, generated using Drosophila RNAi Screening Center Integrative Ortholog Prediction Tool (DIOPT) and Online Mendelian Inheritance of Man (OMIM)-derived known disease/phenotype associations and corresponding MIM numbers. Subset of 31 Drosophila genes co-regulated with *brm, osa,* and/or *Snr1*.

**Supplementary Table S6.13. Percent knockdown of co-regulated genes.** Average RNAi-mediated qRT-PCR knockdown of co-regulated genes.

**Supplementary Table S6.14. Quantification of changes in behavior from knockdown of co-regulated genes.** Quantification of changes in behavior from RNAi knockdown of co-regulated genes. (A) ANOVAs for startle response. (B) Fisher's Exact Tests for tapping behavior. (C) ANOVAs for sleep and activity measurements. Line and Sex are fixed effects. df: degrees of freedom, SS: Type III Sum of Squares, MS: Mean Squares.

*Supplementary Files*

**Supplementary File S6.1. Video of tapping behavior in a male fly with knockdown of *vtd* following a startle response.**

**Supplementary File S6.2. Video of control male fly following a startle response.**

# Appendix F

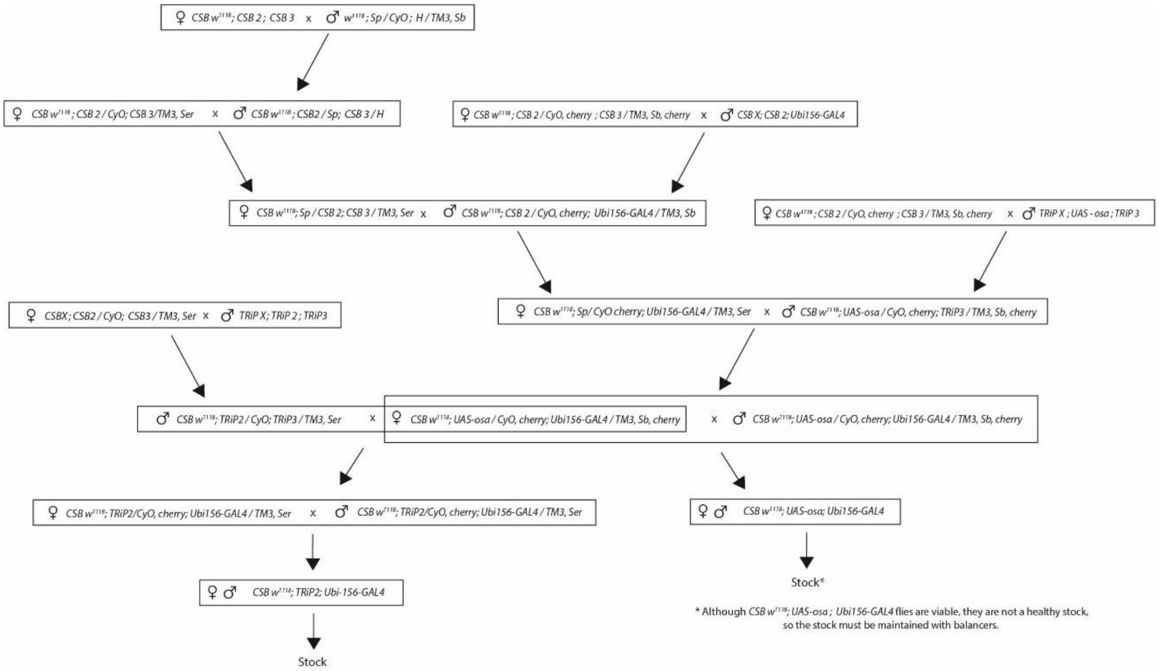
## Supplementary material for Chapter 7

**All files are located at :**

https://clemson.box.com/s/ul80pg6zz382fzsukyxdodvepxi2ccy6

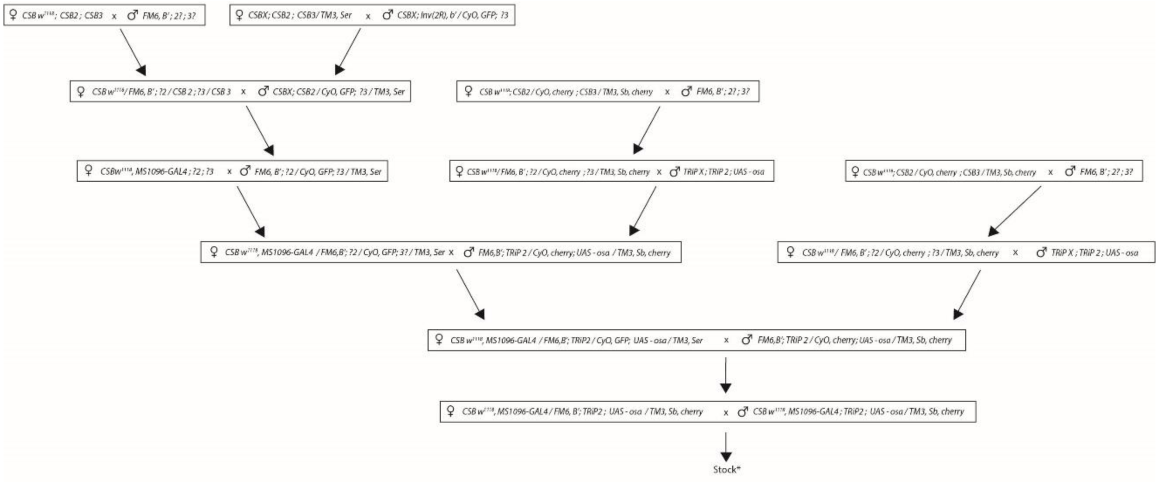
//MacPherson_dissertation/chapter07_ARID1B_modifiers

## *Supplementary Figures*


♀ CSB $w^{1118}$; CSB 2 ; CSB 3 x ♂ $w^{1118}$; Sp / CyO ; H / TM3, Sb
♀ CSB $w^{1118}$ ; CSB 2 / CyO; CSB 3/TM3, Ser x ♂ CSB $w^{1118}$ ; CSB2 / Sp; CSB 3 / H
♀ CSB $w^{1118}$ ; CSB 2 / CyO, cherry ; CSB 3 / TM3, Sb, cherry x ♂ CSB X; CSB 2; Ubi156-GAL4
♀ CSB $w^{1118}$; Sp / CSB 2; CSB 3 / TM3, Ser x ♂ CSB $w^{1118}$; CSB 2 / CyO, cherry; Ubi156-GAL4 / TM3, Sb
♀ CSB $w^{1118}$ ; CSB 2 / CyO, cherry ; CSB 3 / TM3, Sb, cherry x ♂ TRiP X ; UAS - osa ; TRiP 3
♀ CSBX; CSB2 / CyO; CSB3 / TM3, Ser x ♂ TRiP X; TRiP 2 ; TRiP3
♀ CSB $w^{1118}$; Sp/ CyO cherry; Ubi156-GAL4 / TM3, Ser x ♂ CSB $w^{1118}$; UAS-osa / CyO, cherry; TRiP3 / TM3, Sb, cherry
♂ CSB $w^{1118}$; TRiP2 / CyO; TRiP3 / TM3, Ser x ♀ CSB $w^{1118}$; UAS-osa / CyO, cherry; Ubi156-GAL4 / TM3, Sb, cherry x ♂ CSB $w^{1118}$; UAS-osa / CyO, cherry; Ubi156-GAL4 / TM3, Sb, cherry
♀ CSB $w^{1118}$; TRiP2/CyO, cherry; Ubi156-GAL4 / TM3, Ser x ♂ CSB $w^{1118}$; TRiP2/CyO, cherry; Ubi156-GAL4 / TM3, Ser
♀ ♂ CSB $w^{1118}$; UAS-osa; Ubi156-GAL4
Stock*
♀ ♂ CSB $w^{1118}$; TRiP2; Ubi-156-GAL4
Stock

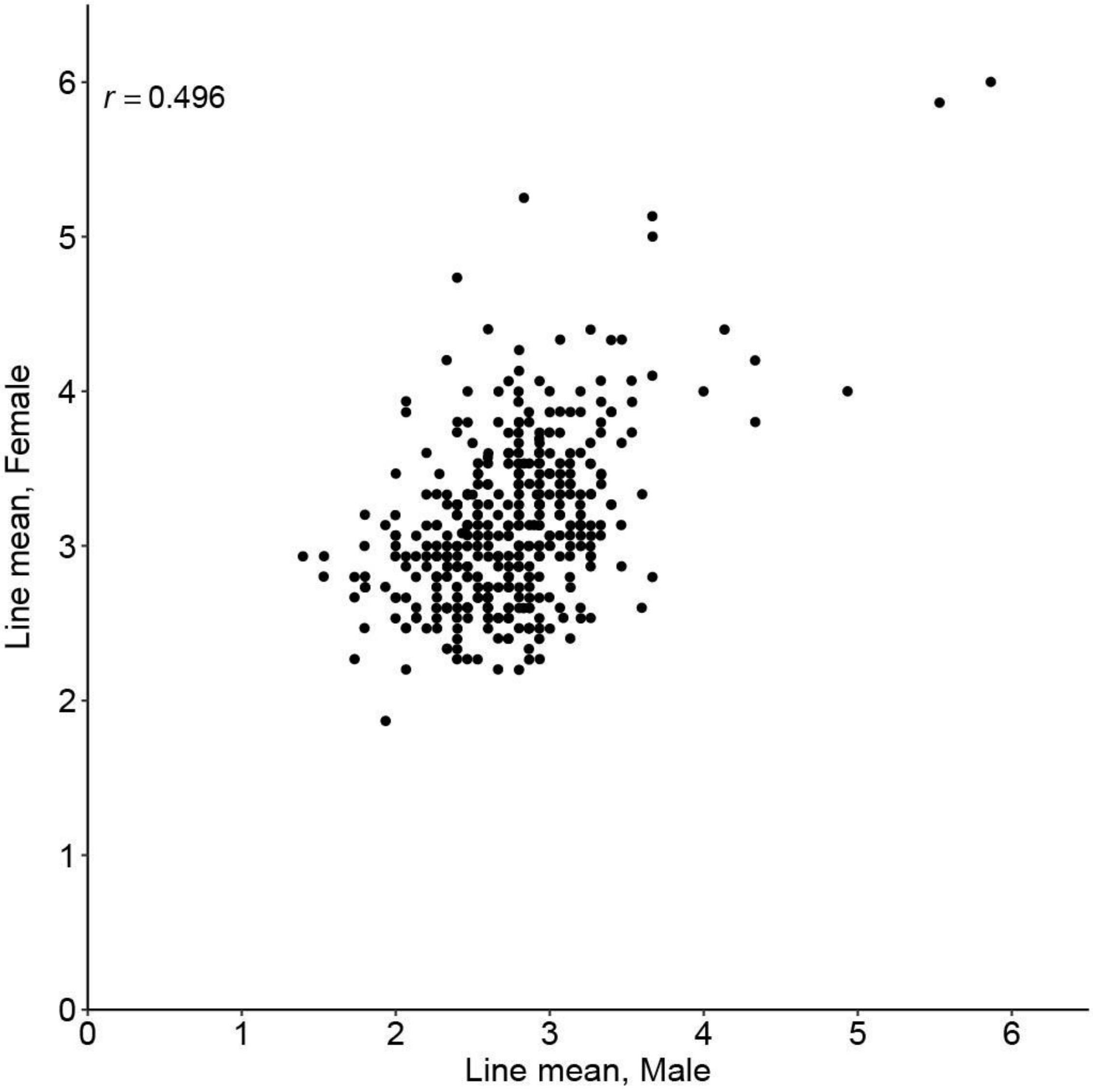
* Although CSB $w^{1118}$; UAS-osa ; Ubi156-GAL4 flies are viable, they are not a healthy stock, so the stock must be maintained with balancers.

**Supplementary Figure S7.1. Crossing scheme to generate *Ubi156>osa* flies.** The TRiP RNAi *attp40 UAS-osa* construct was combined in the same genetic background as the *Ubi156-GAL4* RNAi driver. All females were crossed as virgins. Balancer stock genotypes prior to isogenization on a CantonS-B background can be found in Table S1.


♀ CSB $w^{1118}$; CSB2 ; CSB3 x ♂ FM6, B' ; 2? ; 3?
♀ CSBX; CSB2 ; CSB3/ TM3, Ser x ♂ CSBX; Inv(2R), b' / CyO, GFP; ?3
♀ CSB $w^{1118}$/ FM6, B' ; ?2 / CSB 2 ; ?3 / CSB 3 x ♂ CSBX; CSB2 / CyO, GFP; ?3 / TM3, Ser
♀ CSB $w^{1118}$; CSB2 / CyO, cherry ; CSB3 / TM3, Sb, cherry x ♂ FM6, B' ; 2? ; 3?
♀ CSB$w^{1118}$, MS1096-GAL4 ; ?2 ; ?3 x ♂ FM6, B' ; ?2 / CyO, GFP; ?3 / TM3, Ser
♀ CSB $w^{1118}$/ FM6, B' ; ?2 / CyO, cherry ; ?3 / TM3, Sb, cherry x ♂ TRiP X ; TRiP 2 ; UAS - osa
♀ CSB $w^{1118}$; CSB2 / CyO, cherry ; CSB3 / TM3, Sb, cherry x ♂ FM6, B' ; 2? ; 3?
♀ CSB $w^{1118}$, MS1096-GAL4 / FM6,B'; ?2 / CyO, GFP; 3? / TM3, Ser x ♂ FM6,B'; TRiP 2 / CyO, cherry; UAS - osa / TM3, Sb, cherry
♀ CSB $w^{1118}$/ FM6, B' ; ?2 / CyO, cherry ; ?3 / TM3, Sb, cherry x ♂ TRiP X ; TRiP 2 ; UAS - osa
♀ CSB $w^{1118}$, MS1096-GAL4 / FM6,B'; TRiP2 / CyO, GFP; UAS - osa / TM3, Ser x ♂ FM6,B'; TRiP 2 / CyO, cherry; UAS - osa / TM3, Sb, cherry
♀ CSB $w^{1118}$, MS1096-GAL4 / FM6, B'; TRiP2 ; UAS - osa / TM3, Sb, cherry x ♂ CSB $w^{1118}$, MS1096-GAL4 ; TRiP2 ; UAS - osa / TM3, Sb, cherry
Stock*

* Although *CSB $w^{1118}$, MS1096-GAL4; TRiP2 ; UAS - osa* flies are viable, they are not fertile, so the stock must be maintained with balancers.

**Supplementary Figure S7.2. Crossing scheme to generate *MS1096>osa* flies.** The TRiP RNAi *attp2 UAS-osa* construct was combined in the same genetic background as the *MS1096-GAL4* RNAi driver. All females were crossed as virgins. Balancer stock genotypes prior to isogenization on a CantonS-B background can be found in Table S1.



**Supplementary Figure S7.3. Correlation of line means across sexes.** Line means from male flies plotted against line means from female flies for each of the 392 DGRP lines scored. Pearson's correlation (*r*) is shown on the figure.

*Supplementary Tables*

**Supplementary Table S7.1. Drosophila stock reagents.** Genotype details and stock numbers for lines used and/or generated in this study. (A) non-DGRP Drosophila stocks, (B) DGRP 2.0 and DGRP 3.0 stocks.

**Supplementary Table S7.2. Quantification of percent knockdown.** Quantitative real-time PCR data showing average *Ubi156-GAL4*-mediated percent knockdown of *osa*. BDRC: Bloomington Drosophila Research Center

**Supplementary Table S7.3. Analysis of sleep and activity phenotypes from knockdown of *osa*.** Quantification and statistical analysis of changes in sleep and activity phenotypes from *Ubi156-GAL4*-mediated knockdown of *osa*, using the *attp40* UAS-osa RNAI line. Genotype and Sex are fixed effects, Rep is a random effect. df: degrees of freedom, SS: Type III Sum of Squares, MS: Mean Squares.

**Supplementary Table S4. Analysis of background-dependent changes in sleep and activity phenotypes from knockdown of *osa*.** Quantification and statistical analysis of changes in sleep and activity phenotypes across different DGRP lines from *Ubi156-GAL4*-mediated knockdown of *osa*, using the *attp40* UAS-osa RNiI line. Genotype and Sex are fixed effects, Line and Rep are random effects. df: degrees of freedom, SS: Type III Sum of Squares, MS: Mean Squares.

**Supplementary Table S5. Line means for wing vein scores from knockdown flies.** Line means representing wing vein score for *DGRP X/MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa* females and *MS1096-GAL4; DGRP 2/TRiP 2*; *DGRP 3/UAS-osa* males, or both sexes, for 392 DGRP lines.

**Supplementary Table S6. Vein counts for control flies.** Vein counts for flies without *MS1096-GAL4* driver (*DGRP X/FM6, B'; DGRP 2/TRiP 2; DGRP 3/UAS-osa* females and *DGRP X; DGRP 2/TRiP 2; DGRP 3/UAS-osa* males). "1" and "0" indicates presence or absence of wildtype (intact) wing vein phenotype, respectively. "category" indicates whether line mean for the corresponding RNAi genotype was in the highest, lowest, or middle ~10% of line means, sexes separately.

**Supplementary Table S7. Analyses of variance and broad sense heritability for *MS1096>osa* x DGRP flies.** Statistical analysis of wing vein scores for *MS1096>osa* flies. Sex is a fixed effect, Line is a random effect. df: degrees of freedom, SS: Type III Sum of Squares, MS: Mean Squares, $H^2$ : broad sense heritability, n: average sample size.