Investigating the Molecular Basis of Sex-Biased Synaptic Gene Regulation in *Drosophila melanogaster*

By

Andrew Joseph (AJ) Lonski

A.B. in Neuroscience, Princeton University, 2023

Thesis

Submitted in partial fulfillment of the requirements for the Degree of Master of Science in the Graduate Program of Biotechnology at Brown University

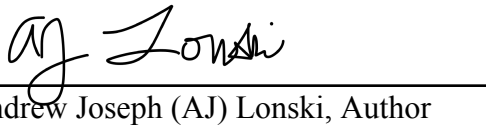
PROVIDENCE, RHODE ISLAND

MAY 2025

AUTHORIZATION TO LEND AND REPRODUCE THE THESIS

As the sole author of this thesis, I authorize Brown University to lend it to other institutions for the purpose of scholarly research.

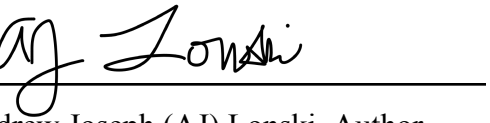
Date 4/30/2025



Andrew Joseph (AJ) Lonski, Author

I further authorize Brown University to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Date 4/30/2025



Andrew Joseph (AJ) Lonski, Author

This thesis by Andrew Joseph (AJ) Lonski is accepted in its present form by the Graduate Program in Biotechnology as satisfying the thesis requirements for the degree of Master of Science.

Date 4/29/2025 Signature: _____

Dr. ERICA LARSCHAN, Advisor

Date 4/29/2025 Signature: K.M. O'Connor-Giles

Dr. KATE O'CONNOR-GILES, Advisor

Date 4/29/2025 Signature: _____

Dr. ALVIN HUANG, Reader

Date 4/29/2025 Signature: _____

Dr. KRISTI WHARTON, Reader

Approved by the Graduate Council

Date _____ Signature: _____

Dr. Thomas Lewis, Dean of the Graduate School

Acknowledgments

I would like to thank my advisors, Erica Larschan and Kate O'Connor-Giles, for introducing me to the world of genetics and genomics. To all other members of my labs — you are the best! Special thanks to the undergrads, Sophie and Liana, for guiding me through various molecular genetics techniques. Thank you to Rajan, Jenn, Scott, and Amanda for helping me whenever the (very) frequent question arose.

Last but not least, thank you to recently minted Dr. James Kentro. It is rare to encounter individuals like James who are so willing to help those around them to excel. Despite the incredible amount on his plate, he consistently and patiently guided me through both the computational and wet-lab aspects of my research. He is the kind of researcher—and person—I aspire to become over the course of my PhD journey at UChicago. I am very thankful to have called him a friend over the last 2 years, and look forward to seeing him thrive as a PI one day.

To my greatest cheerleader and supporter, Byron—I love you so much. Each home cooked meal; each late night snack; your daily kindness and care; they all allowed me to devote myself to my research and passions. Thank you for putting up with me. I am endlessly grateful.

To my family and friends—Dan, Lillie, Jamie, Laura, Philip, Mina, Gawon, Gillian, Kevin, Cody, Luna—love you. Despite our distance, you have consistently done every/ anything to help make my life easier. To the wonderful souls I have met at Brown and in Providence—Sabrina, Esra, Amelia, Kaitlyn, Isabella, Ananya, Jessie, Shevaughn, Sosie, Hannah, Emily, Karma, Sophie, Viv, Tuan, Theo, Nick and Anthony, Rob and Nelson, Jeffrey and Jason, Mark M. and Mark L., Matty and Frank, and many others—thank you for bringing so much joy and brightness to my life.

Table of Contents

Table of Figures

Table of Tables

Abstract of *INVESTIGATING THE MOLECULAR BASIS OF SEX-BIASED SYNAPTIC GENE REGULATION IN DROSOPHILA MELANOGASTER, by AJ LONSKI, ScM, Brown University, May, 2025.*

Motivation: Many higher-order organisms display sexually dimorphic traits, specifically, sex-biased lifespan and aging phenotypes. There are several potential explanations for these phenomena—one of which involves the complex interplay between genes, transcription factors, and epigenetic markers to regulate various cellular and biological processes. Due to the complexity of the brain, sex-biased aging remains poorly understood, especially at the neural level.

Objective: Within this thesis, I had two primary objectives—To characterize sex-biased alternative splice isoforms of synaptic genes across two major *Drosophila melanogaster* developmental stages, and determine whether the temporally coordinated pattern of synaptic gene expression and chromatin accessibility observed across *D. melanogaster* and *M. musculus* (mouse) CNS development is evolutionarily conserved in human brain-derived cells.

Methods: For my first objective, I performed bulk RNA-sequencing analysis and transcriptome alignment on wild-type third instar and adult, male and female *Drosophila*. Next, I identified high-matching human orthologs of several synaptic genes that appear to be coordinately-regulated across both *Drosophila* and mouse development, and then analyzed time-series scRNA-seq and scATAC-seq data of roughly 650,000 human-derived cells.

Results: I identified differential alternative splicing of genes associated with synaptic development and differentially enriched gene ontology pathways between biological sexes and across two major developmental stages. Separately, I found similar patterns of synaptic gene expression observed in flies and mice in several human cerebral organoids, but a unique pattern of chromatin accessibility only present in human cells.

Conclusions: While there appear to be clear sex-biased splicing differences between biological sexes in *D. melanogaster* at the bulk level, future studies should expand splicing analysis across the entire continuum of *Drosophila* development, as well as integrate single-cell multiomics. Interestingly, chromatin accessibility appeared to oppose synaptic gene expression across developmental stages in human cells, suggesting a potential unique mechanism of transcription factor binding and subsequent remodeling of 3D genomic architecture during human development—one that should be investigated using paired multi omics integration. A future study that I have outlined and obtained preliminary results for will characterize the functional consequences of knocking down *onecut*, a candidate transcriptional regulator of synaptic development in *Drosophila*, through in-vivo analysis of the third instar larval neuromuscular junction.

Chapter 1: Introduction

1.1 Sex differences in the brain during aging

Across the animal kingdom, we observe sexual dimorphism in aging, longevity, and the regulation of developmental processes. In several mammalian species, including humans, female lifespan is often significantly longer than that of their male counterparts (Austad and Fischer 2016; Lemaître et al. 2020). The extended biological age of female humans corresponds to specific molecular biomarkers, whose features have been used to predict aging trajectory (Hägg and Jylhävä 2021). In some species, the difference in longevity between sexes is more extreme, as in the case of female short-finned pilot whales, which live twice as long as their male counterparts (Kasuya 1984).

There are several theories as to why we observe these sex-biased differences in lifespan and aging. Notably, the "unguarded X" and "toxic Y" hypotheses argue that differences in lifespan and aging between biological sexes arise due to differences in chromosomes (Connallon et al. 2022). Specifically, a reduced chromosome in the heterogametic sex (i.e., XY sex chromosomes in human males) exposes the other sex chromosome to deleterious, recessive mutations (Bosserman 1985). This exposure, in turn, leads to differences in lifespan, with homogametic species (i.e., XX sex chromosomes in human females) often outliving their heterogametic counterparts (Xirocostas et al. 2020) (Figure 1). Others posit that increased sexual competition within the same sex accelerates senescence due to increased resource investment in mating (Grunst et al. 2018). Disentangling the role of epigenetic alterations in differential aging and lifespan from other biological processes has also provided insight into sexual dimorphism. Steve Horvath, notably, has developed several DNA methylation-based 'epigenetic clocks', which predict biological age with tissue and cell-type specificity and can even generalize across different species (Horvath 2013; Horvath and Raj 2018). Others have used single-nucleotide

(sn)RNA-seq data to predict biological age with cell-type specificity in the fruit fly *Drosophila melanogaster* (Lu et al. 2023). While much work has been done to explore the mechanistic components of sex-biased aging, a robust framework that integrates the various aspects of both healthy and dysfunctional aging processes remains to be developed.

A fascinating aspect of sexual dimorphism in humans is the difference in age-related dysfunction and the manifestation of neurodegenerative diseases. For example, elderly females are significantly more likely to develop Alzheimer's Disease (AD) than elderly males, and this risk continues to increase with age (Niu et al. 2017). At the same time, amyotrophic lateral sclerosis (ALS), a progressive neurodegenerative disease that impacts motor function, has a higher incidence in males than in females (McCombe and Henderson 2010). More confusingly, the symptoms and progression of neurodegeneration may also differ between sexes. The prevalence of multiple sclerosis (MS) is higher in females (Koch-Henriksen and Sørensen 2010; Trojano et al. 2012), but males that develop MS will deteriorate faster (Voskuhl and Gold 2012; Lopez-Lee et al. 2024). The regulatory machinery underlying these sex-based differences in aging and neurodegeneration remains unclear, leading some contemporary researchers to pursue gene regulatory network (GRN) inference.

1.2 Gene regulatory networks (GRNs)

Given the complexity of coordinating tissue-specific and systems-level developmental processes, a potential driver of sex-biased aging is gene regulatory networks. A GRN is a graph-based model that describes how different molecular signals interact with one another to regulate gene expression. At the cellular level, transcriptional machinery, such as enhancers and silencers, work together to temporally regulate transcription factor (TF) binding at mRNA sites (Levine and Davidson 2005). The binding of pioneer TFs to chromatin, which alters chromatin

accessibility by nucleosome displacement, allows other TFs to bind to the newly opened chromatin, thereby increasing (or decreasing) gene transcription by a promoter. Through the binding of different TFs to DNA, combined with other regulatory factors, mRNA (and subsequently protein) can be synthesized from DNA (Figure 2).

In a GRN, the combinatorial interaction between genomic enhancers and TFs may be represented through a larger node-based network structure, allowing for the characterization of large regulatory regions across the genome (Emmert-Streib et al. 2014; Janssens et al. 2022). Effectively, genes/TFs are represented as nodes, while edges are the regulatory interactions between genes (Badia-I-Mompel et al. 2023). This graph-based representation, coupled with computational modeling, can provide meaningful information underlying highly specific gene regulation. For instance, a study of bulk mouse tissue found sex-specific gene expression by cell type, which is mediated by specific GRN interactions (Villa et al. 2018; Lopez-Lee et al. 2024). At the single-cell resolution, GRNs provide a way to interpret how the activity of developmental trajectories (such as neurogenesis) are regulated in cellular subpopulations (Janssens et al. 2022). This approach is able to capture nonlinear chromosomal or long-range gene-gene interactions (Miele and Dekker 2008), and can be integrated with 3D chromatin structure to provide insight into the regulatory role of 3D genome structures (Badia-I-Mompel et al. 2023; Pancaldi 2023). Through GRN inference and subsequent pathway validation, we may be able to better understand observed sex differences in aging.

1.3 Alternative Splicing (AS)

Another regulatory aspect of gene expression is alternative RNA splicing (AS), a process by which a single gene is able to produce multiple different mRNA products (and subsequent protein products). In the AS process, a ribonucleotide complex called the spliceosome removes

the intronic region from pre-mRNA through catalytic cleavage of the 5' splice site (Angarola and Anczuków 2021). The spliceosome, together with regulatory splicing factors and several other types of inputs, generates specific mature mRNA isoforms depending on which exons are skipped. As a result of specific dosage or regulatory inputs, seven main splicing event types may arise: skipping exon (SE), alternative 5'/3' splice sites (A5/A3), mutually exclusive exons (MX), retained intron (RI), and alternative first/last exons (AF/AL) (Blencowe 2006; Wang et al. 2015). These local AS events, in turn, are mediated by epigenetic, transcriptional, and post-translational modifications (PTMs) across development (Angarola and Anczuków 2021).

In humans, 95% of all genes undergo AS, with aberrant splicing machinery contributing to neurodegenerative disease and age-related dysfunction (Pan et al. 2008; Wang et al. 2008; Scotti and Swanson 2016; Deschênes and Chabot 2017; Angarola and Anczuków 2021). For example, depletion of the nuclear-binding protein TDP-43, a pathological hallmark of amyloid lateral sclerosis (ALS) and frontotemporal dementia (FTD), leads to non-conserved cryptic exon splicing, disrupting translation and promoting nonsense-mediated decay (Ling et al. 2015). It is also apparent that many sexually dimorphic traits that are observed in mammalian species may arise due to differences in alternative splicing. A bulk transcriptomic analysis of human brains identified sex-biased exon usage (Kang et al. 2011). Similarly, bulk RNA-seq tissue from the Chinese horseshoe bat brain revealed differentially expressed genes (DEGs) and splice isoforms between sexes, with functional roles in sexually dimorphic physiological processes and RNA splicing regulation, respectively (Chen et al. 2023). While informative, it is difficult to isolate the regulatory impact of an individual gene/transcript in mammalian systems due to longer lifespans and greater tissue diversity, especially at the single-cell resolution (Chen et al. 2019). Therefore, studying the AS isoforms in well-characterized model organisms, like *Drosophila melanogaster*,

at multiple developmental time points could provide better insight into sex-biased biological and epigenetic mechanisms.

1.4 Using *Drosophila melanogaster* as a model organism

Drosophila melanogaster is a powerful model organism to better understand the mechanisms of observed differences between biological sexes across aging and developmental stages, especially with our understanding of the *Drosophila* dosage compensation complex (DCC). While the (sex) chromosomal differences between biological sexes would normally cause a lethal imbalance in gene expression, the DCC addresses this otherwise lethal change through transcriptional regulation of the male X chromosome (Marín et al. 2000; Conrad and Akhtar 2012). The *Drosophila* DCC comprises the proteins MSL1, MSL2, MSL3, MOF, and MLE, and long non-coding RNAs roX1 and roX2 (Kuroda et al. 2016; Samata and Akhtar 2018; Tikhonova et al. 2022). In *Drosophila*, the MSL complex regulates genes on the X chromosome differentially based on the X/A chromosome ratio. The presence of a single X chromosome (in the case of biologically sexed males) prevents translation of the female-specific RNA-binding gene, *Sex lethal* (*Sxl*). Without this expression, a regulatory cascade allows for the targeting of *msl2* mRNA and remodeling of the chromatin landscape through the DCC (Lucchesi and Kuroda 2015). In male flies, the MSL complex of the DCC up-regulates global H4K16ac acetylation—binding chromatin and enhancing gene expression—on the single male X chromosome. In the case of females, the *Sxl* gene expresses SXL, preventing *msl2* mRNA and DCC recruitment.

This differential regulation of sex chromosomes in *Drosophila*, combined with observed differences in lifespan between sexes, suggests that there are genetic and/or epigenetic drivers that function in a sex-specific manner. For instance, the loss of global heterochromatin has been observed across numerous species as a hallmark of aging. A study by Brown et al. revealed

disproportionately greater misexpression on the *Drosophila* Y chromosome during aging and the disappearance of repressive chromatin from pericentromeric regions and the Y chromosome (Brown et al. 2017). After generating XXY females, X0 males, and XYY males, the researchers directly correlated chromatin de-repression and average lifespan with Y chromosome number (Brown et al. 2017), suggesting the existence of sexually dimorphic regulatory mechanisms.

Recent evidence has demonstrated that chromatin-linked adapter for MSL proteins (CLAMP), a zinc-finger pioneer transcription factor, undergoes synergistic binding with the MSL complex, even promoting long-range interactions on the *Drosophila* male X-chromosome (Jordan and Larschan 2021). Interestingly, sex-specific alternative splicing has been shown across the *Drosophila* genome during the period of early embryogenesis. CLAMP-dependent differential expression and gene splicing occur in both the embryonic and third instar larval stages, suggesting CLAMP's context-specific dual role in splicing and transcription (Ray et al. 2023). Moreover, researchers found that CLAMP regulates alternative splicing of the *sxl* transcript and Sxl protein expression in females, reinforcing its role in sex-specific DCC regulation. While important on its own, CLAMP interacts with other regulators, like the pioneer TF Zelda, to activate the *Drosophila* zygotic genome and remodel chromatin accessibility at embryonic promoters (Duan et al. 2021). Additionally, the CLAMP and MLE interaction influences MSL protein binding in *Drosophila* and drives male-specific DCC recruitment (Albig et al. 2019; Tikhonova et al. 2022; Eggers et al. 2023; Tikhonova et al. 2024).

Multimodal genomic analysis has revealed CLAMP and DEAF1 as key repressors of 485 genes important to synaptic development and function that are coordinately-expressed across *Drosophila* developmental stages (Kentro et al. 2024). By genes important to synaptic development and function I refer to scaffolding molecules, synaptic vesicle tethering machinery, cell adhesion molecules, neurotransmitter receptors and transporters, and ion channel

transporters—from henceforth, I will refer to these synaptic effector genes simply as "synaptic genes" in the context of my analysis. While insightful, it is still unclear how this synaptic gene activation and repression are temporally coordinated in *Drosophila*, given the combinatorial interaction of multiple TFs and genes. In investigating the gene regulatory basis of sex-biased aging, researchers have identified several candidate synaptic transcriptional regulators (Bayarmagnai et al. 2012; Leyva-Díaz and Hobert 2022; Kentro et al. 2024).

1.5 Thesis objectives – Understanding the molecular basis for sex-specific synaptic gene regulation during aging

The advent of multi-omics data analysis may provide insight into the genetic and epigenetic regulatory mechanisms that underlie synapse maintenance and, thus, neurodegeneration and aging. Within the scope of this thesis, I pursued two main aims. First, I sought to identify transcript variability and differential alternative splicing of synaptic genes and how these differences differed across sexes and development. To do so, I utilized bulk RNA-seq data from adult and third instar larval *Drosophila melanogaster* brains to identify differential splicing across the genome, and then quantifying how synaptic genes are differentially expressed. By comparing the types of local splicing events across developmental stages and across conditions, I provide a comprehensive characterization of sex-biased synaptic splicing isoforms in *Drosophila*.

A second goal of this project was to characterize temporal aspects of synaptic gene regulation in more advanced model organisms. Kentro et al. found that temporally coordinated synaptic gene expression regulates neuronal connectivity across *Drosophila* development, and this temporal coordination is evolutionarily conserved in mice (Kentro et al. 2024). Due to the conservation of temporally coordinated synaptic gene expression in both *Drosophila* and mice, I

sought to characterize gene expression in human synaptic orthologs across the developmental stages of human cerebral organoid samples at the single-cell resolution, thereby reinforcing the conserved mechanism of temporally coordinated synaptic gene regulation across development in both invertebrate and mammalian species.

At the end of my thesis, I discuss the TF *onecut* as a candidate regulator of synaptic development and maintenance across Drosophila development (Kentro et al. 2024). To better understand the functional role of *onecut* in synaptogenesis and synaptic maintenance in *Drosophila*, I generated a knockdown of *onecut* with RNA interference (RNAi) and conducted in vivo validation. By quantifying the synapses within third instar larval NMJ, future studies determine whether and how *onecut* regulated synaptic bouton growth and morphology.

1.6 Figures

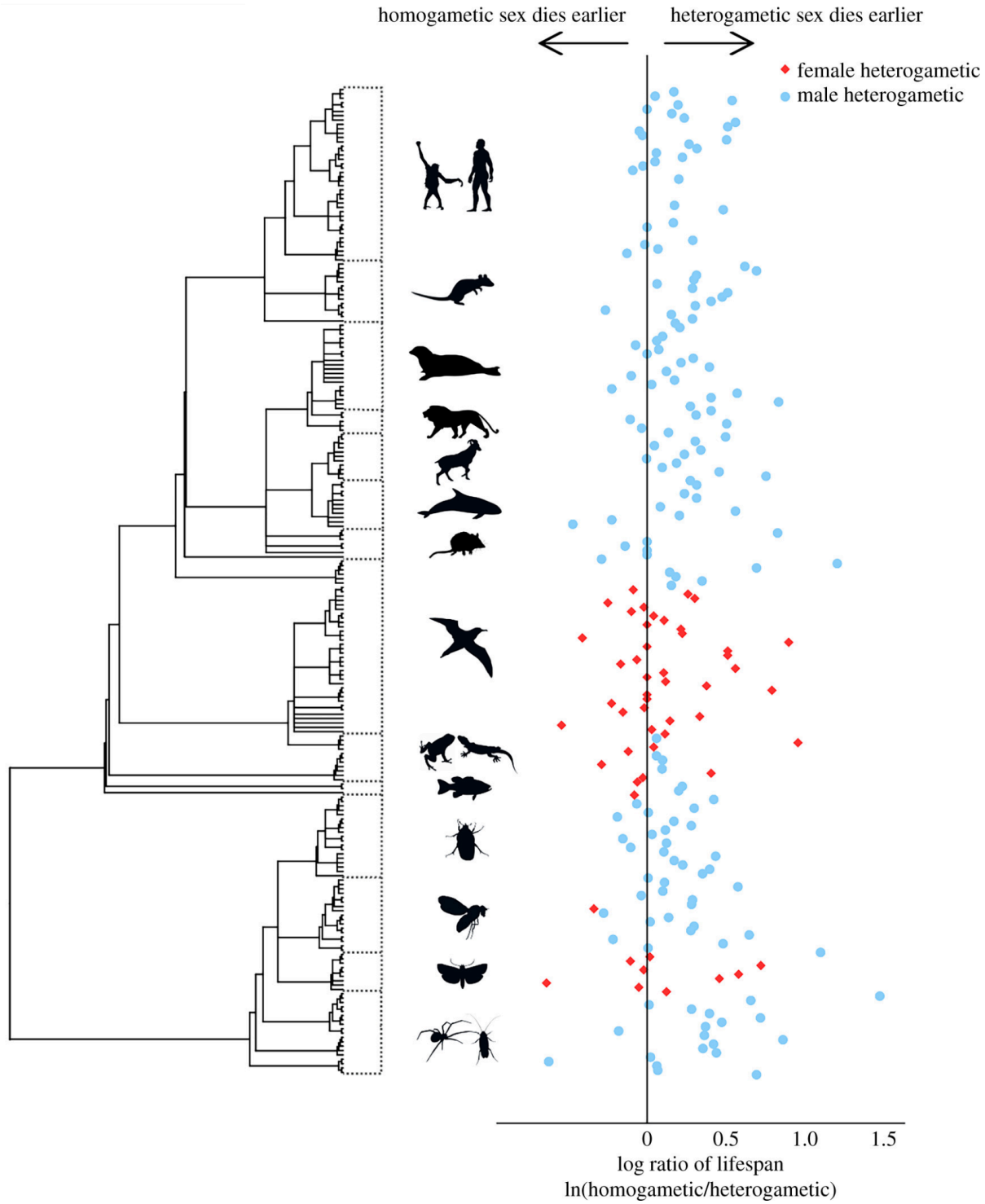


Figure 1: Sexually dimorphic life spans across the animal kingdom. Phylogenetic clustering from primates to insects reveals that male heterogametic sex generally dies earlier than female heterogametic sex in a given species. Adapted from Xirocostas et al. (2020).

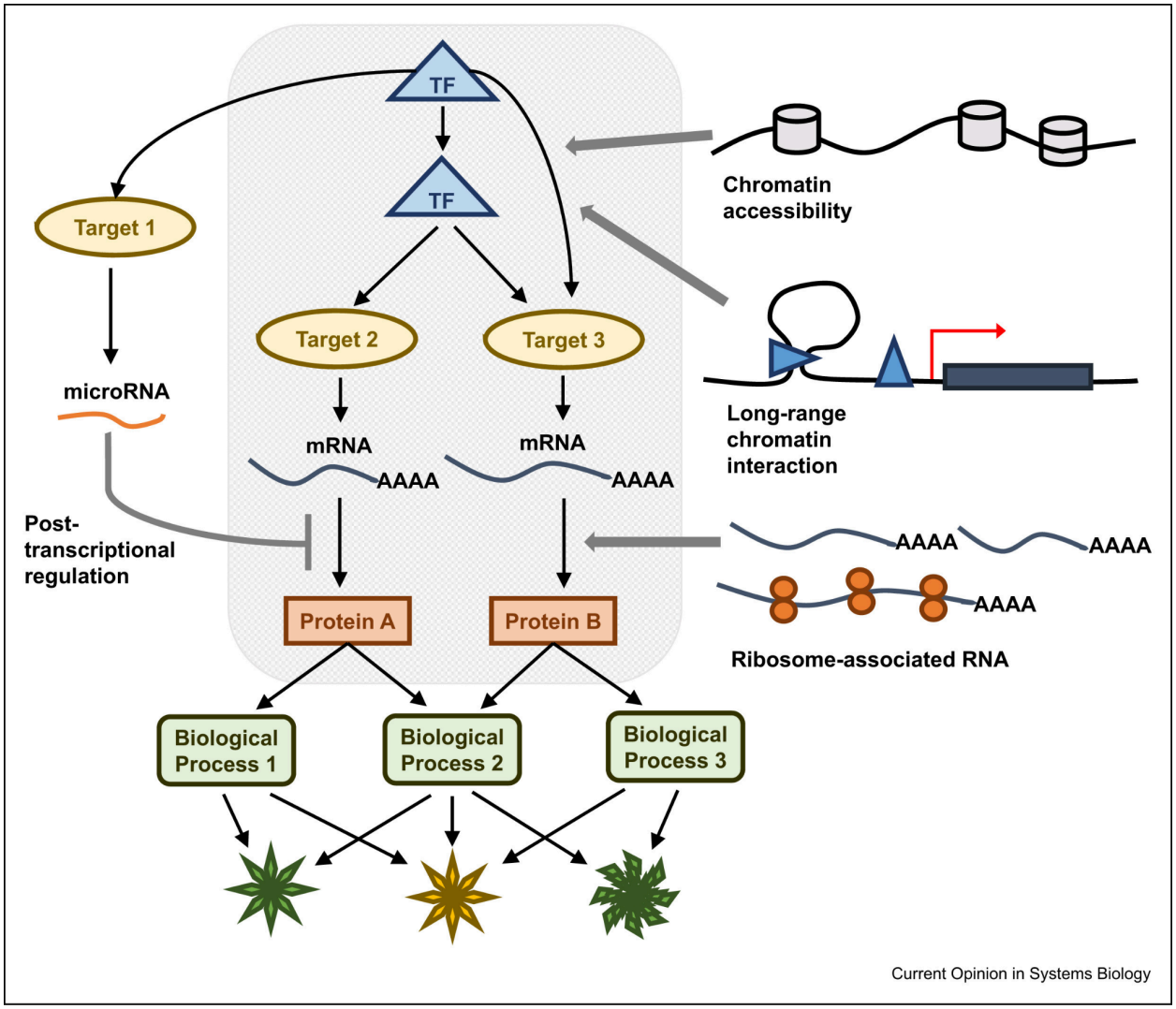


Figure 2: Pathways of a GRN to downstream phenotypes. The gray area contains a canonical GRN with TF-binding to regulate gene expression. Upstream and downstream regulatory pathways of the GRN are indicated by arrow directionality. Adapted from Qian and Huang (2020).

1.7 References

Albig, Christian, Evgeniya Tikhonova, Silke Krause, Oksana Maksimenko, Catherine Regnard, and Peter B Becker. 2019. "Factor Cooperation for Chromosome Discrimination in Drosophila." *Nucleic Acids Research* 47 (4): 1706–24. https://doi.org/10.1093/nar/gky1238.

Angarola, Brittany Lynn, and Olga Anczuków. 2021. "Splicing Alterations in Healthy Aging and Disease." *Wiley Interdisciplinary Reviews. RNA* 12 (4): e1643. https://doi.org/10.1002/wrna.1643.

Austad, Steven N., and Kathleen E. Fischer. 2016. "Sex Differences in Lifespan." *Cell Metabolism* 23 (6): 1022–33. https://doi.org/10.1016/j.cmet.2016.05.019.

Badia-I-Mompel, Pau, Lorna Wessels, Sophia Müller-Dott, Rémi Trimbour, Ricardo O. Ramirez Flores, Ricard Argelaguet, and Julio Saez-Rodriguez. 2023. "Gene Regulatory Network Inference in the Era of Single-Cell Multi-Omics." *Nature Reviews. Genetics* 24 (11): 739–54. https://doi.org/10.1038/s41576-023-00618-5.

Bayarmagnai, Battuya, Brandon N. Nicolay, Abul B. M. M. K. Islam, Nuria Lopez-Bigas, and Maxim V. Frolov. 2012. "Drosophila GAGA Factor Is Required for Full Activation of the dE2f1-Yki/Sd Transcriptional Program." *Cell Cycle (Georgetown, Tex.)* 11 (22): 4191–4202. https://doi.org/10.4161/cc.22486.

Blencowe, Benjamin J. 2006. "Alternative Splicing: New Insights from Global Analyses." *Cell* 126 (1): 37–47. https://doi.org/10.1016/j.cell.2006.06.023.

Brown, Emily J., Alison H. Nguyen, and Doris Bachtrog. 2019. "The Y Chromosome Contributes to Sex-Specific Aging in Drosophila." bioRxiv. https://doi.org/10.1101/156042.

Chen, Geng, Baitang Ning, and Tieliu Shi. 2019. "Single-Cell RNA-Seq Technologies and Related Computational Data Analysis." *Frontiers in Genetics* 10:317. https://doi.org/10.3389/fgene.2019.00317.

Chen, Wenli, Weiwei Zhou, Qianqian Li, and Xiuguang Mao. 2023. “Sex Differences in Gene Expression and Alternative Splicing in the Chinese Horseshoe Bat.” *PeerJ* 11 (April):e15231. https://doi.org/10.7717/peerj.15231.

Connallon, Tim, Isobel J. Beasley, Yasmine McDonough, and Filip Ruzicka. 2022. “How Much Does the Unguarded X Contribute to Sex Differences in Life Span?” *Evolution Letters* 6 (4): 319–29. https://doi.org/10.1002/evl3.292.

Conrad, Thomas, and Asifa Akhtar. 2012. “Dosage Compensation in Drosophila Melanogaster: Epigenetic Fine-Tuning of Chromosome-Wide Transcription.” *Nature Reviews. Genetics* 13 (2): 123–34. https://doi.org/10.1038/nrg3124.

Deschênes, Mathieu, and Benoit Chabot. 2017. “The Emerging Role of Alternative Splicing in Senescence and Aging.” *Aging Cell* 16 (5): 918–33. https://doi.org/10.1111/acel.12646.

Duan, Jingyue, Leila Rieder, Megan M Colonnetta, Annie Huang, Mary Mckenney, Scott Watters, Girish Deshpande, William Jordan, Nicolas Fawzi, and Erica Larschan. n.d. “CLAMP and Zelda Function Together to Promote Drosophila Zygotic Genome Activation.” *eLife* 10:e69937. https://doi.org/10.7554/eLife.69937.

Eggers, Nikolas, Fotios Gkountromichos, Silke Krause, Aline Campos-Sparr, and Peter B Becker. 2023. “Physical Interaction between MSL2 and CLAMP Assures Direct Cooperativity and Prevents Competition at Composite Binding Sites.” *Nucleic Acids Research* 51 (17): 9039–54. https://doi.org/10.1093/nar/gkad680.

Emmert-Streib, Frank, Matthias Dehmer, and Benjamin Haibe-Kains. 2014. “Gene Regulatory Networks and Their Applications: Understanding Biological and Medical Problems in Terms of Networks.” *Frontiers in Cell and Developmental Biology* 2 (August):38. https://doi.org/10.3389/fcell.2014.00038.

Grunst, Melissa L., Andrea S. Grunst, Vincent A. Formica, Marisa L. Korody, Adam M. Betuel, Margarida Barcelo-Serra, Rusty A. Gonser, and Elaina M. Tuttle. 2018. "Actuarial Senescence in a Dimorphic Bird: Different Rates of Ageing in Morphs with Discrete Reproductive Strategies." *Proceedings of the Royal Society B: Biological Sciences* 285 (1892): 20182053. https://doi.org/10.1098/rspb.2018.2053.

Hägg, Sara, and Juulia Jylhävä. 2021. "Sex Differences in Biological Aging with a Focus on Human Studies." *eLife* 10 (May):e63425. https://doi.org/10.7554/eLife.63425.

Horvath, Steve. 2013. "DNA Methylation Age of Human Tissues and Cell Types." *Genome Biology* 14 (10): R115. https://doi.org/10.1186/gb-2013-14-10-r115.

Horvath, Steve, and Kenneth Raj. 2018. "DNA Methylation-Based Biomarkers and the Epigenetic Clock Theory of Ageing." *Nature Reviews. Genetics* 19 (6): 371–84. https://doi.org/10.1038/s41576-018-0004-3.

Janssens, Jasper, Sara Aibar, Ibrahim Ihsan Taskiran, Joy N. Ismail, Alicia Estacio Gomez, Gabriel Aughey, Katina I. Spanier, et al. 2022. "Decoding Gene Regulation in the Fly Brain." *Nature* 601 (7894): 630–36. https://doi.org/10.1038/s41586-021-04262-z.

Jordan, William, and Erica Larschan. 2021. "The Zinc Finger Protein CLAMP Promotes Long-Range Chromatin Interactions That Mediate Dosage Compensation of the Drosophila Male X-Chromosome." *Epigenetics & Chromatin* 14 (June):29. https://doi.org/10.1186/s13072-021-00399-3.

Kang, Hyo Jung, Yuka Imamura Kawasawa, Feng Cheng, Ying Zhu, Xuming Xu, Mingfeng Li, André M. M. Sousa, et al. 2011. "Spatiotemporal Transcriptome of the Human Brain." *Nature* 478 (7370): 483–89. https://doi.org/10.1038/nature10523.

KASUYA, T. 1984. "Age Determination and Growth of the Short-Finned Pilot Whale off the Pacific Coast of Japan." *Sci. Rep. Whales Res. Inst.* 35:57–91.

Kentro, James A., Gunjan Singh, Tuan M. Pham, Justin Currie, Saniya Khullar, Audrey T. Medeiros, Maria Tsiarli, Erica Larschan, and Kate M. O'Connor-Giles. 2025. "Conserved Transcription Factors Coordinate Synaptic Gene Expression through Repression." *bioRxiv*, February, 2024.10.30.621128. https://doi.org/10.1101/2024.10.30.621128.

Koch-Henriksen, Nils, and Per Soelberg Sørensen. 2010. "The Changing Demographic Pattern of Multiple Sclerosis Epidemiology." *The Lancet. Neurology* 9 (5): 520–32. https://doi.org/10.1016/S1474-4422(10)70064-8.

Kuroda, Mitzi I., Andres Hilfiker, and John C. Lucchesi. 2016. "Dosage Compensation in Drosophila—a Model for the Coordinate Regulation of Transcription." *Genetics* 204 (2): 435–50. https://doi.org/10.1534/genetics.115.185108.

Lemaître, Jean-François, Victor Ronget, Morgane Tidière, Dominique Allainé, Vérane Berger, Aurélie Cohas, Fernando Colchero, et al. 2020. "Sex Differences in Adult Lifespan and Aging Rates of Mortality across Wild Mammals." *Proceedings of the National Academy of Sciences of the United States of America* 117 (15): 8546–53. https://doi.org/10.1073/pnas.1911999117.

Levine, Michael, and Eric H. Davidson. 2005. "Gene Regulatory Networks for Development." *Proceedings of the National Academy of Sciences of the United States of America* 102 (14): 4936–42. https://doi.org/10.1073/pnas.0408031102.

Leyva-Díaz, Eduardo, and Oliver Hobert. 2022. "Robust Regulatory Architecture of Pan-Neuronal Gene Expression." *Current Biology : CB* 32 (8): 1715-1727.e8. https://doi.org/10.1016/j.cub.2022.02.040.

Ling, Jonathan P., Olga Pletnikova, Juan C. Troncoso, and Philip C. Wong. 2015. "TDP-43 Repression of Nonconserved Cryptic Exons Is Compromised in ALS-FTD." *Science (New York, N.Y.)* 349 (6248): 650–55. https://doi.org/10.1126/science.aab0983.

Lopez-Lee, Chloe, Eileen Ruth S. Torres, Gillian Carling, and Li Gan. 2024. "Mechanisms of Sex Differences in Alzheimer's Disease." *Neuron* 112 (8): 1208–21. https://doi.org/10.1016/j.neuron.2024.01.024.

Lu, Tzu-Chiao, Maria Brbić, Ye-Jin Park, Tyler Jackson, Jiaye Chen, Sai Saroja Kolluru, Yanyan Qi, et al. 2023. "Aging Fly Cell Atlas Identifies Exhaustive Aging Features at Cellular Resolution." *Science (New York, N.Y.)* 380 (6650): eadg0934. https://doi.org/10.1126/science.adg0934.

Lucchesi, John C., and Mitzi I. Kuroda. 2015. "Dosage Compensation in Drosophila." *Cold Spring Harbor Perspectives in Biology* 7 (5): a019398. https://doi.org/10.1101/cshperspect.a019398.

Marín, I., M. L. Siegal, and B. S. Baker. 2000. "The Evolution of Dosage-Compensation Mechanisms." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 22 (12): 1106–14. https://doi.org/10.1002/1521-1878(200012)22:12<1106::AID-BIES8>3.0.CO;2-W.

McCombe, Pamela A., and Robert D. Henderson. 2010. "Effects of Gender in Amyotrophic Lateral Sclerosis." *Gender Medicine* 7 (6): 557–70. https://doi.org/10.1016/j.genm.2010.11.010.

Miele, Adriana, and Job Dekker. 2008. "Long-Range Chromosomal Interactions and Gene Regulation." *Molecular bioSystems* 4 (11): 1046–57. https://doi.org/10.1039/b803580f.

Niu, H., I. Álvarez-Álvarez, F. Guillén-Grima, and I. Aguinaga-Ontoso. 2017. "Prevalence and Incidence of Alzheimer's Disease in Europe: A Meta-Analysis." *Neurología (English Edition)* 32 (8): 523–32. https://doi.org/10.1016/j.nrleng.2016.02.009.

Pan, Qun, Ofer Shai, Leo J. Lee, Brendan J. Frey, and Benjamin J. Blencowe. 2008. "Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by

High-Throughput Sequencing." *Nature Genetics* 40 (12): 1413–15. https://doi.org/10.1038/ng.259.

Pancaldi, Vera. 2023. "Network Models of Chromatin Structure." *Current Opinion in Genetics & Development* 80 (June):102051. https://doi.org/10.1016/j.gde.2023.102051.

Qian, Yichun, and Shao-shan Carol Huang. 2020. "Improving Plant Gene Regulatory Network Inference by Integrative Analysis of Multi-Omics and High Resolution Data Sets." *Current Opinion in Systems Biology* 22 (August):8–15. https://doi.org/10.1016/j.coisb.2020.07.010.

"R. Trivers: Social Evolution. Menlo Park, California: The Benjamin/Cummings Publishing Company, Inc., 1985 - Bosserman - 1985 - Behavioral Science - Wiley Online Library." n.d. Accessed April 13, 2025. https://onlinelibrary.wiley.com/doi/10.1002/bs.3830300409.

Ray, Mukulika, Ashley Mae Conard, Jennifer Urban, Pranav Mahableshwarkar, Joseph Aguilera, Annie Huang, Smriti Vaidyanathan, and Erica Larschan. 2023. "Sex-Specific Splicing Occurs Genome-Wide during Early Drosophila Embryogenesis." *eLife* 12 (July):e87865. https://doi.org/10.7554/eLife.87865.

Samata, Maria, and Asifa Akhtar. 2018. "Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs." *Annual Review of Biochemistry* 87 (June):323–50. https://doi.org/10.1146/annurev-biochem-062917-011816.

Scotti, Marina M., and Maurice S. Swanson. 2016. "RNA Mis-Splicing in Disease." *Nature Reviews. Genetics* 17 (1): 19–32. https://doi.org/10.1038/nrg.2015.3.

Tikhonova, Evgeniya, Sofia Mariasina, Olga Arkova, Oksana Maksimenko, Pavel Georgiev, and Artem Bonchuk. 2022. "Dimerization Activity of a Disordered N-Terminal Domain from Drosophila CLAMP Protein." *International Journal of Molecular Sciences* 23 (7): 3862. https://doi.org/10.3390/ijms23073862.

Tikhonova, Evgeniya, Sofia Mariasina, Sergey Efimov, Vladimir Polshakov, Oksana Maksimenko, Pavel Georgiev, and Artem Bonchuk. 2022. “Structural Basis for Interaction between CLAMP and MSL2 Proteins Involved in the Specific Recruitment of the Dosage Compensation Complex in Drosophila.” *Nucleic Acids Research* 50 (11): 6521–31. https://doi.org/10.1093/nar/gkac455.

Tikhonova, Evgeniya, Anastasia Revel-Muroz, Pavel Georgiev, and Oksana Maksimenko. 2024. “Interaction of MLE with CLAMP Zinc Finger Is Involved in Proper MSL Proteins Binding to Chromosomes in Drosophila.” *Open Biology* 14 (3): 230270. https://doi.org/10.1098/rsob.230270.

Trojano, Maria, Guglielmo Lucchese, Giusi Graziano, Bruce V. Taylor, Steve Simpson, Vito Lepore, Francois Grand’Maison, et al. 2012. “Geographical Variations in Sex Ratio Trends over Time in Multiple Sclerosis.” *PLoS ONE* 7 (10): e48078. https://doi.org/10.1371/journal.pone.0048078.

Villa, Alessandro, Paolo Gelosa, Laura Castiglioni, Mauro Cimino, Nicoletta Rizzi, Giovanna Pepe, Federica Lolli, et al. 2018. “Sex-Specific Features of Microglia from Adult Mice.” *Cell Reports* 23 (12): 3501–11. https://doi.org/10.1016/j.celrep.2018.05.048.

Voskuhl, Rhonda R., and Stefan M. Gold. 2012. “Sex-Related Factors in Multiple Sclerosis: Genetic, Hormonal and Environmental Contributions.” *Nature Reviews. Neurology* 8 (5): 255–63. https://doi.org/10.1038/nrneurol.2012.43.

Wang, Eric T., Rickard Sandberg, Shujun Luo, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore, Gary P. Schroth, and Christopher B. Burge. 2008. “Alternative Isoform Regulation in Human Tissue Transcriptomes.” *Nature* 456 (7221): 470–76. https://doi.org/10.1038/nature07509.

WANG, YAN, JING LIU, BO HUANG, YAN-MEI XU, JING LI, LIN-FENG HUANG, JIN LIN, et al. 2015. “Mechanism of Alternative Splicing and Its Regulation.” *Biomedical Reports* 3 (2): 152–58. https://doi.org/10.3892/br.2014.407.

Xirocostas, Zoe A., Susan E. Everingham, and Angela T. Moles. 2020. “The Sex with the Reduced Sex Chromosome Dies Earlier: A Comparison across the Tree of Life.” *Biology Letters* 16 (3): 20190867. https://doi.org/10.1098/rsbl.2019.0867.

Chapter 2: Differential alternative splicing of synaptic genes across development and biological sex in *Drosophila melanogaster*

2.1 Introduction

A fascinating biological mechanism of many eukaryotic organisms is the ability for a single gene to encode several different proteins. The process by which this mechanism is achieved is called alternative splicing (AS), where a single gene produces several distinct splice isoforms through different exon-joining combinations. These splice isoforms, or mRNA transcripts, in turn, can produce distinct protein structures through translation. In humans, AS impacts 95% of genes (Pan et al., 2008), and errors in this splicing process appear to mediate neurodegeneration and abnormal aging (Wang et al., 2008; Scotti & Swanson, 2016; Deschênes and Chabot, 2017; Angarola and Anczuków, 2021).

Interestingly, sex-biased gene expression and alternative splicing have been observed in humans and several other mammalian species (Kang et al., 2011; Villa et al., 2018; Chen et al., 2023; Lopez-Lee et al., 2024). The characterization of splicing behavior across developmental time points may reveal an important function for temporally coordinated splicing in sex-biased aging and development. Specifically, I aimed to investigate and quantify splicing differences between *D. melanogaster* males and females, from third instar larvae and adult flies, which may provide insight into sex-biased synaptic gene dysregulation during the process of aging.

2.2 Results

2.2.1 Characterizing global differential alternative splicing events

To establish a broad view of sex and developmental-specific splicing differences, I began by quantifying genome-wide differential alternative splicing (AS) events from the bulk tissue of *D. melanogaster* male and female third instar larvae and adults. I aligned the bulk data using

Spliced Transcripts Alignment to a Reference (STAR), given its ability to precisely detect splice junctions (Dobin et al. 2013). Following alignment, I used RSEM to quantify transcript per million (TPM), SUPPA to first convert TPM to percent spliced in (PSI) for each type of local AS event, and then computed the differential PSI (ΔPSI) between sexes and across development. I defined genome-wide "significant" differential splicing events with $|\Delta PSI| > 0.2$ and $p < 0.05$.

I observed 1822 differential splicing events between third instar and adult males, with mutually exclusive (MX) AS occurring most frequently, and 433 events in females. Between sexes, third instar males and females demonstrated 58 differential AS events, while adult male and female flies demonstrated a greater amount of differential AS events at 93, indicating an increase of AS events with age. These results confirmed both global sex- and development-specific splicing events, prompting me to investigate whether differential splicing occurs in synaptic genes.

2.2.2 *Drosophila melanogaster* undergo sex-specific differential splicing of synaptic genes

After my genome-wide differential splicing analysis, I sought to identify and quantify the differential splicing of synaptic isoforms across sexes and developmental stages. To investigate this, I focused on a subset of synaptic genes that were previously observed to be temporally regulated across *D.* melanogaster development—I filtered the genome-wide dPSI data for 1765 transcripts generated using a list of 485 synaptic genes (Kentro et al., 2024). I initially used the same genome-wide significance thresholds of $|\Delta PSI| > 0.2$ & p-value < 0.05 to quantify synaptic dPSI, but I relaxed the threshold to $|\Delta PSI| > 0.1$ & $p < 0.05$ to identify more synaptic AS events for downstream analysis.

I found 195 differentially spliced synaptic isoforms from third instar to adult males, while only 56 synaptic AS events between third instar and adult females (Figure 3). The observed

difference in synaptic AS events between biological sexes appeared to reflect the genome-wide reduction in AS for females. In comparing biological sexes, I found 2 differential AS events between third instar males and females, with upregulation of both events in male larvae: I observed an SE event for CG42268 and an A5 event for dpr3. I observed six differential synaptic AS events between adult males and females (Table 1, Figure 4). Unlike the third instar larval stage, adult males and females each underwent upregulated splicing of three genes.

Interestingly, nAChRα7 appears to undergo the same RI event type in adult males and females but at different coordinates on the X chromosome: 19328554-19329933 (female-upregulated) and 19347638-19348447 (male-upregulated). Due to this regional difference, I then mapped the genomic coordinates of nAChRα7 to its unique mRNA transcript, identifying the female-upregulated transcript as FBtr0301835 (nAChRα7-RE). Since the genomic coordinates of the male-upregulated AS event differed, I found that either FBtr0301835 (nAChRα7-RE) or FBtr0074680 (nAChRα7-RA) could be transcribed. The observed differential splicing events of synaptic genes, particularly nAChRalpha7 and dpr3, encouraged me to explore the functional consequences of various splice isoforms (i.e., the male-upregulated transcript nAChRα7-RA and female-upregulated transcript nAChRα7-RE). I next examined the biological roles of these differentially spliced genes by performing gene ontology (GO) enrichment analysis.

2.2.3 *Drosophila* display shared and conserved GO enrichment of differential synaptic AS events

After identifying several unique synaptic splice isoforms across conditions, I next sought to identify GO annotation across the three major sub-ontologies: Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF). To do so, I performed

Benjamini-Hochberg (BH) adjustment (Benjamini and Hochberg, 1995) across all genes with significant differential AS events and then filtered for the list of differentially spliced synaptic events identified in the previous section. As expected, I identified differentially enriched GO pathways between third instar and adult males, as well as third instar and adult females (Figure S1). Generally, the most significant GO annotations were involved in synaptic processes (adj. p.value < 0.001). Intuitively, the observation of many synaptic processes makes sense due to the reconfiguration of the CNS when moving from the third instar larval stage to adulthood. I additionally constructed heat maps of the unique splice isoforms within each enriched pathway and their corresponding gene. When comparing third instar to adult males, I found that *cpx*, *nSyb*, *Rim,* and *Syt1* splice isoforms were present across the most enrichment pathways (Figure S1). The enrichment from these four genes also spanned the most GO categories in the third instar to adult female differential splicing comparison.

I also observed differential GO enrichment when comparing between sexes at each developmental timepoint. For the third instar male vs female comparison, I identified six differential GO categories extending from the gene *dpr3*. To visualize this functional enrichment analysis, I generated a gene-concept network that plots the linkage between identified genes and GO enrichment terms (Figure 5). In the adult male vs female comparison, 51 differential GO annotations were observed from the 6 synaptic differential AS events identified in section 2.2.2. I generated a gene-concept network of the top 20 most enriched GO terms and 6 synaptic genes, revealing differentially enriched pathways for many different types of synaptic and sensory processes (Figure 6). Interestingly, I found that several of the top 20 GO terms were related to vision, such as *detection of light stimulus*, *phototransduction*, *response to light stimulus,* and *visual behavior*. While valuable in providing pathway information, I remained unclear as to how

these GO terms grouped into broader pathway categories—leading me to pursue semantic similarity analysis.

2.2.4 GO semantic similarity analysis using hierarchical clustering reveals distinct parental synaptic pathways

Due to GO term enrichment occurring across a diverse set of pathways, I became interested in identifying larger pathways that encompassed the total differential GO annotations between adult males and females. To further synthesize the functional relevance of the synaptic AS events, I implemented Gene Ontology Semantic Similarity Analysis (GOSemSim), allowing me to cluster GO terms into larger, more interpretable categories. GOSemSim is a method that computes semantic similarity between GO terms through information content (IC) or graph-based methods (Gan, 2014; Yu, 2020). Each of the three major sub-ontology groups (BP, MF, CC) are represented as directed acyclic graphs (DAGs), where nodes represent GO terms and edges between nodes represent a parent-child relationship. Prior comparative analysis of GO semantic similarity measurements has revealed best-use cases for each metric type (Cheng et al., 2019; Xue et al., 2023), leading me to investigate several different implementations. I computed a semantic similarity matrix across all nodes, testing three different similarity metrics: the Jaccard correlation coefficient (JC), IC-based Resnik similarity, and graph-based Wang similarity (see section 2.4.3). Taking these similarity matrices, I then implemented the hierarchical clustering algorithm ward D2, to produce a functional grouping tree diagram of five overarching GO categories spanning all annotations. As only 6 GO terms were identified from differential splicing events in the third instar male vs female comparison, I limited this hierarchical clustering method to the adult male vs female comparison.

I initially used JC to construct a hierarchical tree diagram combining the top 30 GO terms (ranked by BH-adjusted p-value) for all three GO sub-ontologies (BP, MF, CC) (Figure 7). My

hierarchical clustering analysis revealed larger categories of *spine postsynaptic synapse potential, amine metabolic catabolic process, neuron projection membrane, and C-activating component carbon dioxide*. Interestingly, response to light stimulus was placed in its own semantic category, despite the other visual-related pathways clustering within the pathway *C-activating component carbon dioxide*. As comparison for this JC-based hierarchical clustering, I implemented the Resnik and Wang methods, which are GO sub-ontology-specific—requiring me to develop tree diagrams for each of the BP, MF and CC pathways. For BP, Resnik and Wang identified slightly different groups than one another, but collectively differed from JC (Figure 8).

After confirming the overlap in BP-specific GO categorization using Resnik and Wang, I asked whether the two methods co-correlated, and how they differed from the JC similarity matrix. I performed principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation projection (UMAP), comparing the three methods (Figure S2). PCA indeed confirmed that Resnik and Wang captured similar low-dimensional representations of the data that were distinct from JC, with PC1 and PC2 capturing 45% of the variance of the data. Running t-SNE and UMAP with the first 2 components revealed similar findings. After visualization, I computed the Pearson and Spearman correlations between the similarity matrices of the JC, Resnik and Wang methods (Table 2). The Resnik-Wang methods had a Pearson correlation of 0.9131 and a Spearman correlation of 0.8696, once again reinforcing their classification similarity. On the other hand, Resnik-JC had a Pearson/Spearman correlation of 0.2354/0.2654, and Wang-JC had a Pearson/Spearman correlation of 0.1895/0.2517. This distinction of Resnik and Wang from JC, as well as their specific development for GO semantic similarity, suggest that their identified clusters possess greater biological interpretability.

I also performed the same clustering analysis workflow for the CC and MF pathways, once again identifying similarities between Resnik and Wang—both distinct from the JC similarity matrix. Interestingly, Resnik and Wang correlated far less with each other for CC (Spearman correlation = 0.3461) and MF (Spearman correlation = 0.3696) than in BP (Table 2). While the overall correlation between Resnik and Wang were much lower for the CC and MF pathways, there were several potential explanations for this observation. Firstly, BP contained a much higher proportion of identified GO terms than the MF pathway. Prior research found the Resnik metric to perform best for protein-protein interaction (PPI) network BP annotation clustering (Xue et al., 2023), in line with my own BP pathway enrichment observations using differentially spliced transcripts. Compared to BP, the CC and MF pathways contain demonstrably less DAG depth for *D. melanogaster* (Buza et al., 2008). While Resnik relies on IC and considers MICA, Wang relies on structural proximity within a given DAG (see section 2.4.3). Given this discrepancy in DAG representation and subsequent less common ancestors for terms, it is plausible that Resnik and Wang would assign CC and MF-associated terms differently, as I observed.

2.2.5 Synaptic transcripts and GO Terms are uniquely enriched across sexes and developmental periods

To build upon the clustering method of the enriched GO pathways, I next asked whether the synaptic transcripts and their corresponding GO terms were unique to each developmental- or sex-based differential splicing comparison. In doing so, I could better clarify the specific context in which certain types of synaptic AS events occur. I quantified the shared and unique synaptic transcripts across each comparison, defining “unique” isoforms as those that were significantly differentially spliced only in their respective comparison. I identified 42 unique splice isoforms

for the male larval vs adult comparison, 19 isoforms for the female larval vs adult comparison, 3 terms for the adult male vs. female comparison, and no unique transcripts for the third instar vs female comparison (Figure 9A). For adult males vs. females, I identified Fbtr0301835 (nAChRα7-RE), Fbtr0074680 (nAChRα7-RA), and Fbtr0082549 (dpr15-RA) as unique transcripts, with nAChRα7 and dpr15 once again appearing as in prior sections.

In addition to finding unique synaptic transcripts across ¾ conditions, I also quantified the unique synaptic GO annotations per comparison. As in the case above, I "defined" unique GO annotations as those only present within their respective comparison. I identified 87 unique GO annotations for the male larval vs adult comparison, 11 unique terms for the female larval vs adult comparison, 3 unique terms for the third instar male vs. female comparison—and interestingly—45 unique terms for the adult male vs female comparison (Figure 9B). Given the large difference in significant differential synaptic AS events between the developmental stages (larval vs adult females and adult males vs. females), I expected the adult male vs. female comparison to have far fewer annotated GO terms. To better visualize the unique GO terms in each condition, I generated tree maps using REVIGO (Supek et al., 2011) for each sub-ontology, where the log fold change of a term (enrichment) corresponds to its area size. After removing redundant or obsolete GO terms, as well as condensing several GO terms into their parental pathway, I was left with three distinct tree maps (Figure 10). Within BP, I found unique enrichment of several signaling and metabolic pathways: *phospholipase C-activating G protein-coupled receptor signaling pathway, amine metabolic process, cellular biogenic amine catabolic process, ammonium ion metabolic process* (Fig. 10A). Within CC, I identified several receptors and signaling complexes, like *plasma membrane signaling receptor complex, GTPase complex* and (synaptic) *receptor complex* (Fig. 10B). Finally, MF contained terms from different regulatory and binding domains (Fig. 10C).

When considered together, these comparative analyses revealed that synaptic splice isoforms and their corresponding GO pathways are uniquely enriched, dependent on both biological sex and stage of development. The disproportionately greater number of unique GO annotations within the adult male vs. female comparison suggests that differences in synaptic AS become exacerbated as *Drosophila* develop in a sex-specific manner. In turn, these results highlighted a potential regulatory role of AS in sex-biased neuronal development and aging.

2.3 Discussion

In this chapter, I profiled AS of global and synaptic genes across developmental time points and between biological sexes in *Drosophila melanogaster*. I identified distinct patterns of differentially spliced synaptic isoforms and enriched GO pathways across all comparisons. Notably, nAChRα7-RE, nAChRα7-RA, and dpr15-RA emerged as uniquely spliced synaptic isoforms in the adult male vs. female condition.

Alternative splicing impacts almost all of the genome, with a crucial role in promoting molecular diversity. Dysfunctional AS mechanisms have been implicated in several sex-biased neurodegenerative and neurodevelopmental disorders. In both mice and humans, the X chromosome is enriched for neuronal genes, with aberrant enrichment linked to neurological issues, aging-related pathology, and neurodegeneration (Davis et al., 2021). While the mechanism of dosage compensation is evolutionarily conserved across mammalian and invertebrate species—the regulatory impact of the DCC in the adult brain remains poorly understood. Given this uncertainty and the sexually dimorphic AS patterns observed in this chapter, investigating alternative splicing in *Drosophila* serves as a powerful model for determining how AS may contribute to neuronal dysregulation in the aging brain.

The nicotinic acetylcholine receptor subunit α7 (*nAChRα7*) is a ligand-gated ion channel with a significant role in acetylcholine-gated monoatomic cation-selective transport and, therefore, cholinergic signaling as a whole. In humans, its encoding gene CHRNA7 has a prominent role in the cholinergic anti-inflammatory pathway (Pohanka, 2012), with roles in pre- and postsynaptic neuronal signaling processes, dendritic development, and synaptic plasticity (Broide and Leslie, 1999; Fayuk and Yakel, 2007; Wallace and Porter, 2011). Post-mortem analysis of the human brain has revealed altered nAChRα7 function in the hippocampus of schizophrenia and AD patients (Freedman et al., 1995; Guan et al., 2001), and differential expression of various nAChR subunits has been reported in in the peripheral cells of AD and Dementia with Lewy bodies (DLB) patients (Costantini et al., 2023). Interestingly, increased cerebral nAChRα7 distribution across several major brain regions is associated with healthy aging in humans, as shown through a positron emission tomography (PET) study (Coughlin et al., 2018), underscoring the regulatory role of nAChRα7 over the course of development and aging.

Beyond human studies, the dysregulation of nAChRα7 is associated with accelerated cognitive decline in AD mouse models and reduced expression in Parkinson's disease dementia (PDD) mice (Pan et al., 2020; Izuo et al., 2023). In *Drosophila*, recent studies of AD models have found that the human A□ peptide inhibits astrocytic nAChRα7, thereby impairing memory formation by preventing H_2O_2 synthesis (Rabah et al., 2025). By altering this astrocyte-to-neuron signaling, neuronal protection from reactive oxygen species (ROS)-related damage becomes inhibited, impairing long-term memory formation in the mushroom body (Rabah et al., 2023; Rabah et al., 2025). Furthermore, in mice, conditional deletion of nAChRα7 in GABAergic interneurons resulted in male-sex-specific decreased postnatal neurogenesis,

social behaviors, and memory, emphasizing the role of the receptor in sex-specific brain development (Otto et al., 2020).

To my knowledge, no prior studies have investigated the role of aberrant nAChRα7 splicing and transcript isoforms—specifically nAChRα7-RE and nAChRα7-RA—in the context of development and aging. Although quite structurally similar, the two transcript isoforms exhibit distinct splicing differences: the nAChRα7-RE variant is 4214 base pairs (bp) and contains 14 exons (13 coding exons), encoding the longest protein sequence of the four nAChRα7 variants (564 amino acids), while nAChRα7-RA has a length of 4145 bp and contains 15 exons (14 coding exons) to encode a protein of 560 amino acids (Figure 11). In the ventral lateral neurons of *Drosophila* larvae, the nAchR subunits of nAChRα1 and nAChRα6 appear to temporally coordinate cholinergic postsynaptic structures (Rosenthal et al., 2021). Based on this observation, there may be a similar sex-specific role for nAChRα7 for synaptic development and maintenance. Elucidating the regulatory mechanisms that drive nAChRα7 splice variants may shed light on how sex-specific AS contributes to synaptic development and maintenance, as well as its function in driving differential aging trajectories.

These findings raise an additional question—are such sex-biased and developmentally unique synaptic events conserved in more complex organisms? In the next chapter, I explore the gene expression and chromatin accessibility dynamics of human orthologs of these synaptic genes in brain-derived cerebral organoids.

2.4 Materials and Methods

2.4.1 RNA isolation, library preparation, and sequencing

D. melanogaster third instar larvae were sexed, and their brains (optic lobes and VNC) were dissected out in PBS in rapid succession. The larval brains (N = 12) were then immediately

placed in RNase-free 1.5ml tubes (containing 100µl of TRIzol lysis reagent, Invitrogen, USA) on ice. There were four biological replicates (tubes) for each genotype, with each replicate/tube containing 12 brains. The larval brains in each tube were homogenized with a pestle. QIAGEN RNeasy Plus Universal Mini Kit was used for the RNA extraction, and ZYMO RNA Clean & Concentrator-25 kit was used for further purification. The RNA was eluted and confirmed for purity and concentration via 260/280 and 260/230 ratios measured using a Thermo Scientific NanoDrop One^{C}. Illumina libraries were prepared and sequenced at Novogene Corporation for 150bp Poly-A selected mRNA and paired-end sequenced using an Illumina NovaSeq 6000, targeting at least 20 million reads per sample.

Five-day-old *D. melanogaster* adult flies were sexed, and their brains were dissected out in PBS in rapid succession. The adult brains (N = 25) were then immediately placed in RNase-free 1.5ml tubes (containing 100µl of TRIzol lysis reagent, Invitrogen, USA) on ice. There were four biological replicates (tubes for each genotype), with each replicate/tube containing 25 brains. The adult brains in each tube were homogenized with a pestle. QIAGEN RNeasy Plus Universal Mini Kit was used for the RNA extraction, and ZYMO RNA Clean & Concentrator-25 kit was used for further purification. The RNA was eluted and confirmed for purity and concentration via 260/280 and 260/230 ratios measured using a Thermo Scientific NanoDrop One^{C}. Illumina libraries were prepared and sequenced at Novogene Corporation for 150bp Poly-A selected mRNA and paired-end sequenced using an Illumina NovaSeq 6000, targeting at least 40 million reads per sample.

2.4.2 Bulk RNA-seq transcriptomics

Bulk RNA-seq data preprocessing

Raw sequencing data were run for quality check by using FASTQ/0.11.9 (Andrews, 2010). Adapted content was trimmed with TrimGalore/0.6.10 (Krueger et al., 2023). Spliced Transcripts Alignment to a Reference (*STAR*) (Dobin et al., 2013) was used for paired-end read alignment to a *Drosophila* melanogaster reference genome (BDGP6.46.dna.toplevel.fa) and annotation (BDGP6.46.111.gtf) (Öztürk-Çolak et al., 2024). *STAR* implements a suffix array (SA) to identify maximum mappable prefixes (MMPs), matching the longest sequence possible against a reference genome. The unmapped read portion is then matched to the reference genome separately, clustered, and stitched with the first mapped alignment. In doing so, *STAR* finds precise splice junction locations without *a priori* knowledge of splice loci (Dobin et al., 2013), leading me to use *STAR* over pseudoalignment methods like *kallisto* (Bray et al., 2016).

Prior to alignment, a *STAR* genome index was generated with default settings, besides --sjdbOverhang 99. Next, the following parameters were used for read alignment for each processed sample (unlisted parameters were default): --runThreadN 8 --outSAMtype BAM SortedByCoordinate --outSAMattributes NH HI AS NM MD --quantMode TranscriptomeSAM GeneCounts. Samtools/1.16.1 (Danecek et al. 2021) was used to sort samples by left-most alignment coordinates (-@ 32) and index for faster random access with default settings. To check the *STAR* alignment QC metrics, I generated flagstat and idxstats reports for each indexed and sorted BAM file. From Subread/2.0.2, featureCounts (Liao et al. 2014) was used to count BAM reads at gene features with the following parameters: -T 8 -g gene_id -t exon -p.

Quantifying alternative splicing events

RNA-Seq by Expectation Maximization (RSEM) (Li & Dewey, 2011) was used on sorted BAM files to quantify transcript per million (TPM). I began by generating an RSEM reference from my existing *STAR* reference (unlisted parameters were default): --star-path

/path/to/STAR_index --num-threads 32. After RSEM reference generation, TPM quantification was calculated with the following parameters: --paired-end --alignments --no-bam-output --estimate-rspd --num-threads 60.

Next, SUPPA2 (Trincado et al., 2018) was used to generate seven types of genome-wide local AS events for each gene from the input annotation gtf file: generateEvents -f ioe -e SE SS MX RI FL. Running this command generated a .ioe file, with each row corresponding to a unique event ID: <gene_id>;<event-type>:<seqname>:<coordinates-of-the-event>:<strand>.

SUPPA2 was then used for conversion from transcript per million (TPM) to percent spliced in (PSI) for each condition and AS event by running psiPerEvent with default settings. To identify differential splicing of AS events and differential transcript usage, I used the SUPPA2 command: diffSplice -m empirical. Finally, OPTICS density-based clustering analysis (Ankerst et al., 1999) was performed using the differential PSI (dPSI) and transcript usage of local events between conditions: clusterEvents --dpsi <dpsi-file> --psivec <psivec-file> --sig-threshold 0.05 --eps 0.2 --separation 0.11 -dt 0.2 --min-pts 10 -c OPTICS. The overall bulk RNA-seq transcriptomics workflow may be viewed in Figure S3.

2.4.3 Data analysis

Differential synaptic isoform downstream analysis

After AS event quantification, .dpsi files of each AS event type were read into R (4.2.2), and filtered for changes in $|\Delta PSI| \geq 0.2$ across each type of comparison: third instar males vs adult males, third instar females vs adult females, third instar males vs third instar females, adult males vs adult females, third instar male/female vs adult male/female and third instar/adult male vs third instar/adult female. Next, biomaRt (Durinck et al., 2009) was used to map 485 synaptic genes (Kentro et al., 2024) to their corresponding transcript IDs using the FlyBase database in

Ensembl (Harrison et al., 2024). The genome-wide .dpsi files were filtered for synaptic transcripts using the mapped biomaRt output. Finally, the generated SUPPA .ioe files for each type of splicing event were combined, and the genomic coordinates of each event were extracted for each condition. The R package dplyr (Wickham et al., 2023) was used for filtering and sorting to identify the unique vs. shared isoforms between each comparison.

Pathway enrichment analysis

To identify pathways that were differentially enriched across conditions, clusterProfiler (Xu et al., 2024) was used to perform gene ontology (GO) annotation on differentially spliced isoforms across all conditions to generate enrichGO objects, with the following settings: OrgDb = org.Dm.eg.db, keyType = "ENSEMBLTRANS", ont = "ALL", pAdjustMethod = "BH", readable = TRUE. With these settings, Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) were all considered for ontology, as well as Benjamini-Hochberg to compute false discovery rate (FDR) and adjusted p.value < 0.05. This annotation procedure was also used for each AS event type individually (within each of the four differential splicing comparisons). The enrichGO object was filtered for NA rows, sorted by descending adjusted p-values., and only the top 30 enriched terms.

To visualize differentially enriched pathways, ggplot() was first used to generate a dotplot across all combined AS events. Next, cnetplot from clusterProfiler (Wu et al., 2021) was used to visualize the network between input genes/transcripts and the top 20 annotated GO pathways in a circular layout: showCategory = 20, circular = TRUE, cex_label_gene = 1, cex_label_category = 1, colorEdge = TRUE.

GO semantic similarity analysis

Following cnetplot, I used pairwise_termsin() to compute the similarity matrix between all nodes combined across BP, CC, and MF sub-ontologies using the Jaccard correlation coefficient (Jaccard, 1902) with default settings. The similarity matrix was used as input for heatplot() from clusterProfiler to show the most representative genes in the top 20 enriched GO pathway (showCategory = 20, other settings default). As comparison against JC, I used Gene Ontology Semantic Similarity Analysis (GOSemSim) (Yu et al., 2020) to compute semantic similarity among all annotated GO terms and gene products/clusters for each GO sub-ontology pathway using the Resnik and Wang methods. The Resnik method (Resnik, 1999) is an IC-based approach, where the frequency of a GO term t is represented as:

$$p(t) = \frac{n_{t'}}{N} \mid t' \in \{t,\ children\ of\ t\}$$

Here, $n_{t'}$ is the number of the GO term t' and N is all terms in the GO corpus. Using this, we may define IC as: $IC(t) = -log(p(t))$,

where the semantic similarity of GO terms is the negative log probability of a GO term occurring across the entire GO corpus, coupled with the term's most informative common ancestor (MICA). Thus, the Resnik method is defined as the following:

$$sim_{Resnik}(t_1,\ t_2) = IC(MICA)\ ,$$

where $t_1$and t_2 are two distinct GO terms. As opposed to Resnik, the Wang method (Wang et al., 2007) represents semantic similarity through a graph-based analysis, where terms closer to term A in DAG_Ahave greater contribution to A's semantic value. In other words, a GO term t has an S-value defined as its semantic contribution to term A. Therefore, for any t in DAG_A, $S_A(t)$ is defined by the piecewise function:

$$\begin{cases} S_A(A) = 1 \end{cases}$$

$$S_A(t) = max\{w_e \times S_A(t') \mid t' \in children\ of\ (t)\}\ if\ t \neq A \quad ,$$

Here, t' is the child term of t, which is linked for edge $e \in E_A$ by the semantic contribution factor w_e. The semantic value of the term A, $SV(A)$, is then: $SV(A) = \sum_{t \in T_A} S_A(t)$,

With GO terms A and B, Wang semantic similarity is therefore

$$sim_{Wang}(A, B) = \frac{\sum_{t \in T_A \cap T_B} S_A(t) + S_B(t)}{SV(A) + SV(B)} ,$$

where $S_A(t)$ and $S_B(t)$ are the S-values of term t for terms A and B, respectively. Taking the similarity matrices generated by JC, Resnik and Wang methods, I utilized the hierarchical clustering algorithm Ward D2 (hclust_method = "ward.D2", other settings default) to compute the sum of the squared Euclidean distance between data point clusters, minimizing the objective function: $D(c_1, c_2) = \delta^2(c_1, c_2) = \frac{|c_1||c_2|}{|c_1| + |c_2|} ||c_1 - c_2||^2$.

In this equation, $|c_1|$ and $|c_2|$ are the element numbers in clusters c_1 and c_2 , respectively. The squared Euclidean distance between cluster centroids is $||c_1 - c_2||^2$ (for further details, see Ward, 1963; Kaufman and Rousseeuw, 1990; Murtagh and Legendre, 2011; Legendre and Legendre, 2012).

I utilized three dimensionality reduction techniques to compare the first two components of the JC, Resnik and Wang method similarity matrices: principal component analysis (PCA) (Abdi and Williams, 2010), t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten and Hinton, 2008; Cieslak et al., 2020), and uniform manifold approximation and projection (UMAP) (McInnes et al., 2018) given prior successful comparative analysis using

bulk transcriptomic datasets (Yang et al., 2021). The details of UMAP are further discussed in section 3.4.1.

2.4.4 Code availability & resources

All downstream data analysis was conducted in R (4.2.2), and custom scripts are available upon request. All R scripts were executed using an integrated RStudio Server with 64 cores and 492 GB of memory.

2.5 Figures

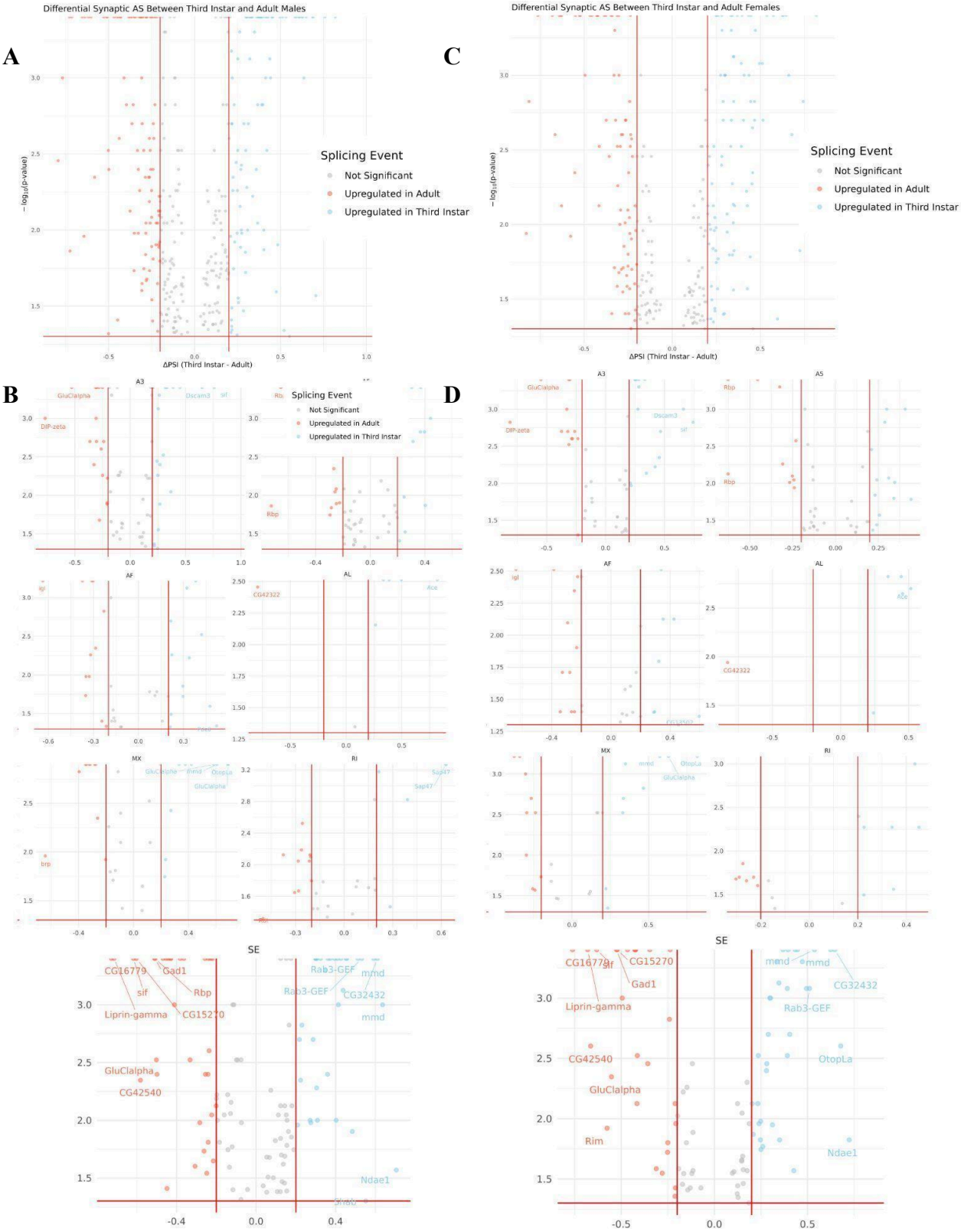

A
Differential Synaptic AS Between Third Instar and Adult Males
Splicing Event
Not Significant
Upregulated in Adult
Upregulated in Third Instar
ΔPSI (Third Instar - Adult)
B
C
Differential Synaptic AS Between Third Instar and Adult Females
D
A3
A5
AF
AL
MX
RI
SE
GluClalpha
Dscam3
sif
DIP-zeta
Rbp
CG42322
Ace
mmd
OtopLa
Sap47
brp
CG16779
Gad1
Rab3-GEF
CG32432
Liprin-gamma
CG15270
CG42540
Rim
Ndae1

Figure 3: Differential splicing of synaptic genes across developmental stages. (A) Volcano plots of all differential synaptic AS events between third instar and adult males. **(B)** Volcano plots of each local AS event type for differentially spliced synaptic genes between third instar and adult males. **(C)** Volcano plots of all differential synaptic AS events between third instar and adult females. **(D)** Volcano plots of each local AS event type for differentially spliced synaptic genes between third instar and adult females. Red colored points represent adult-upregulated splicing, blue colored points represent third instar-upregulated splicing, while gray points represent non-significant splicing events. SE - skipping exon, A5/A3 - alternative 5'/3' splice sites, MX - mutually exclusive exons, RI - retained intron, AF/AL - alternative first/last exons.

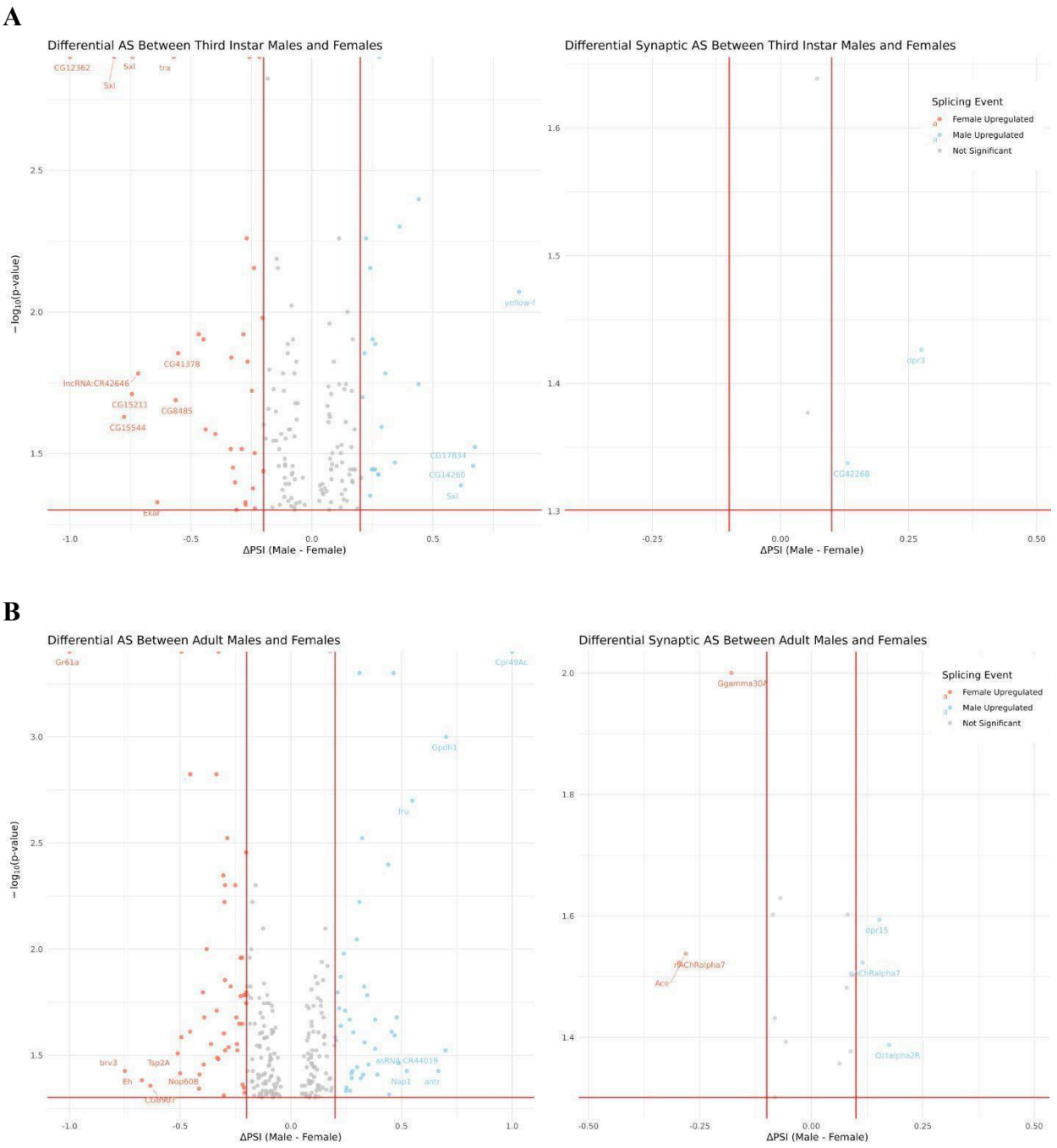


Figure 4: Differential splicing of synaptic genes across sexes. (A) Volcano plots of significant different splicing events between third instar males and females across the genome (left) and of synaptic genes (right). **(B)** Volcano plots of significant differential splicing events between adult males and females across the genome (left) and of synaptic genes (right). Red colored points represent female-upregulated splicing, blue colored points represent male-upregulated splicing, while gray points represent non-significant splicing events.

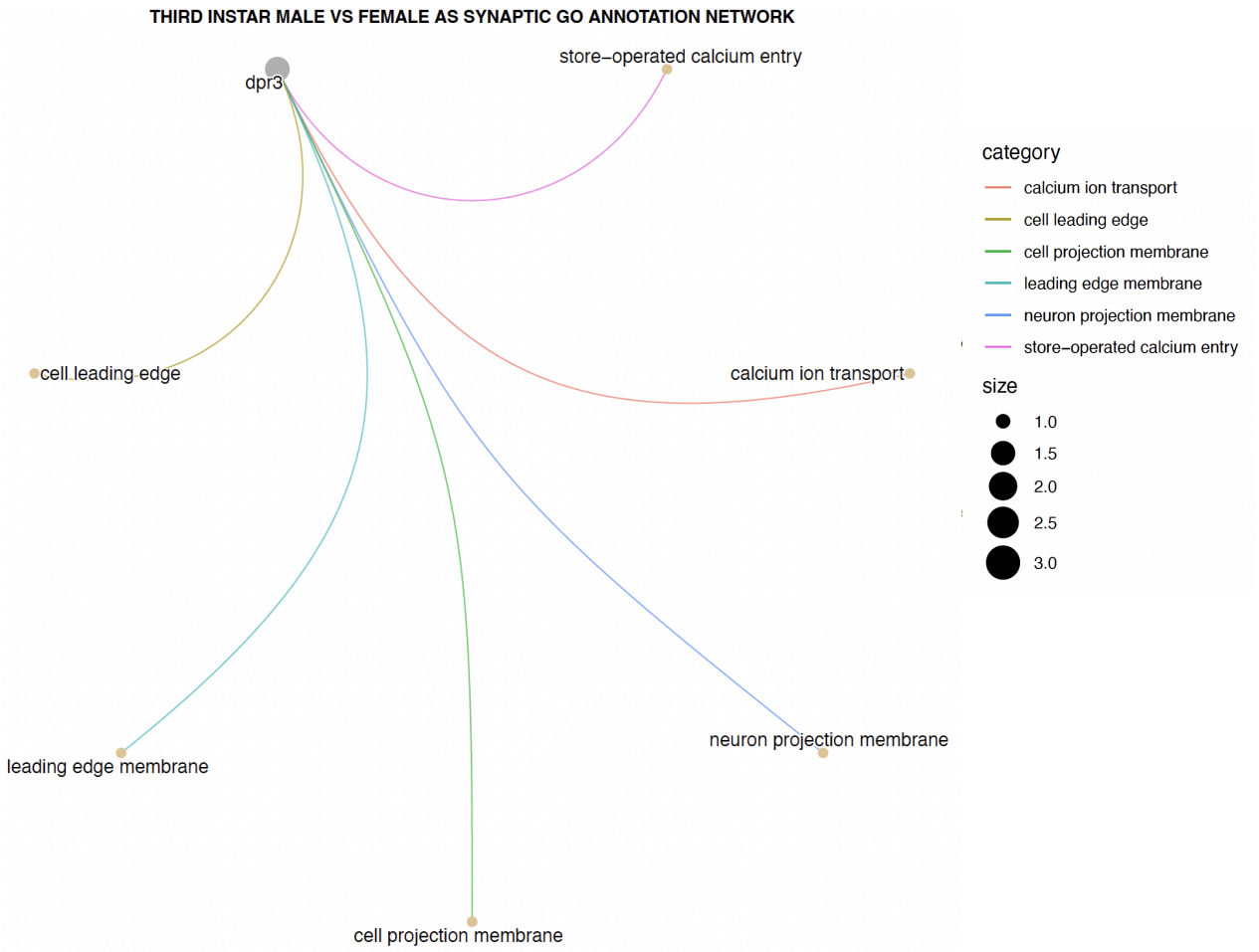


Figure 5: Third instar male vs female gene-concept network of differential synaptic AS events. GO categories are distinguished by different color tracks, while node size represents differentially spliced AS isoforms.

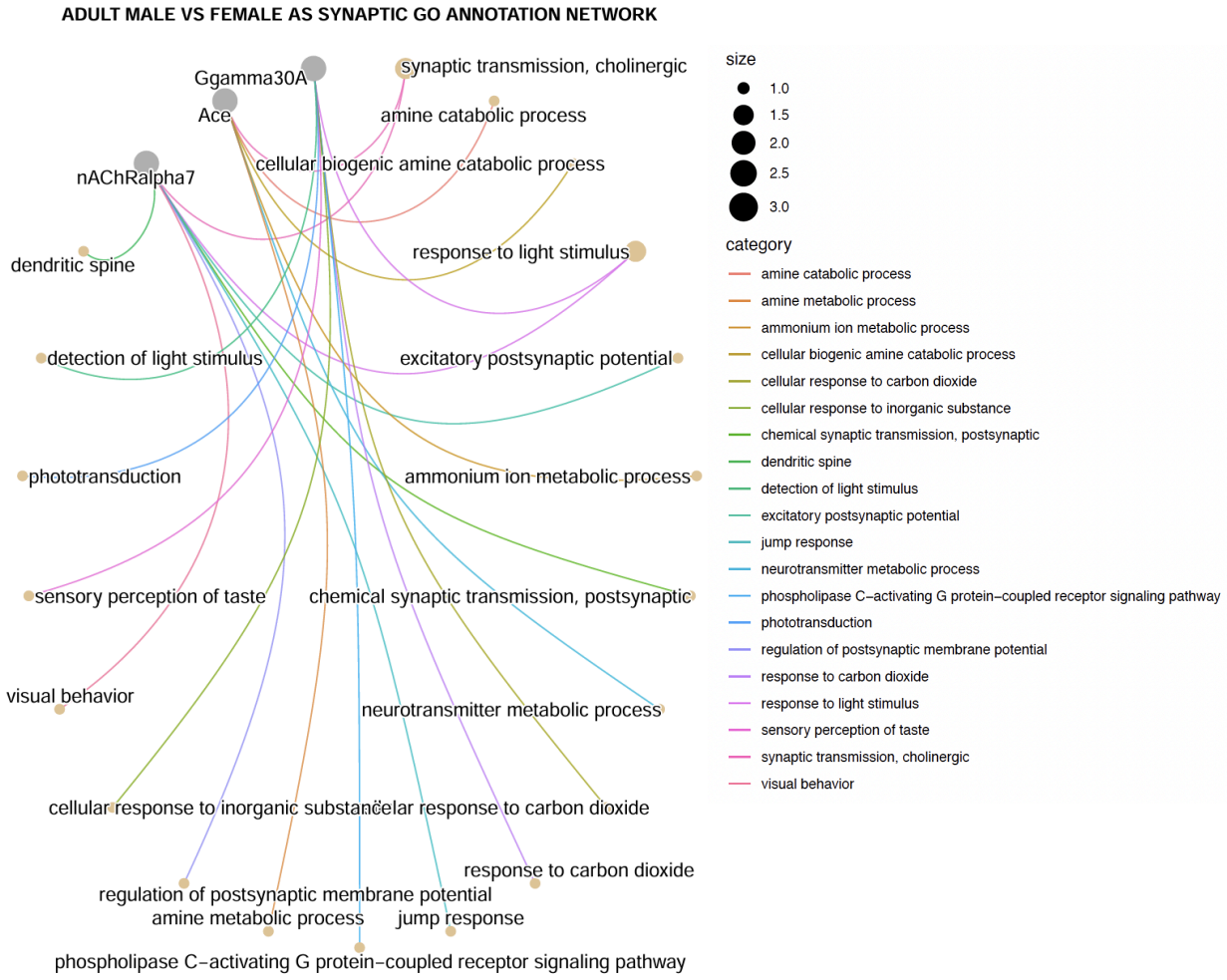


Figure 6: Adult male vs female gene-concept network of differential synaptic AS events. GO categories are distinguished by different color tracks, while node size represents the number of unique differential AS isoforms.

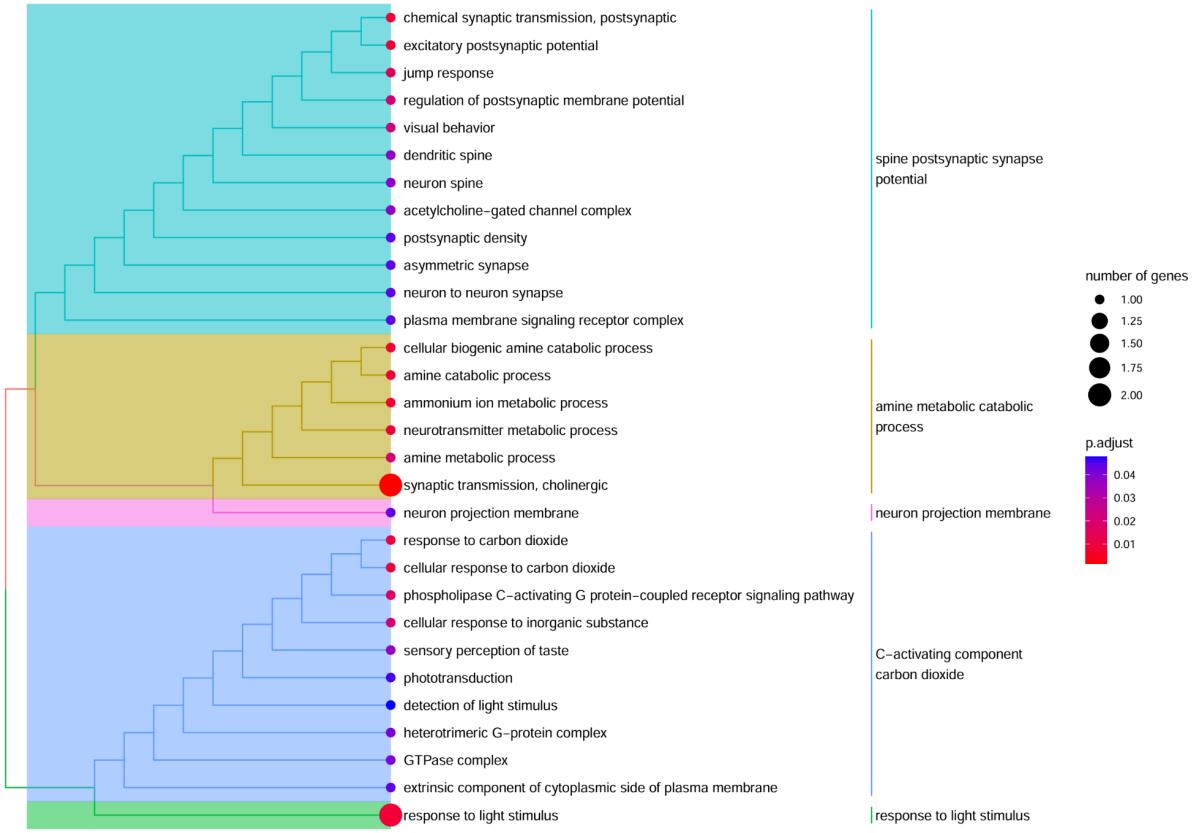
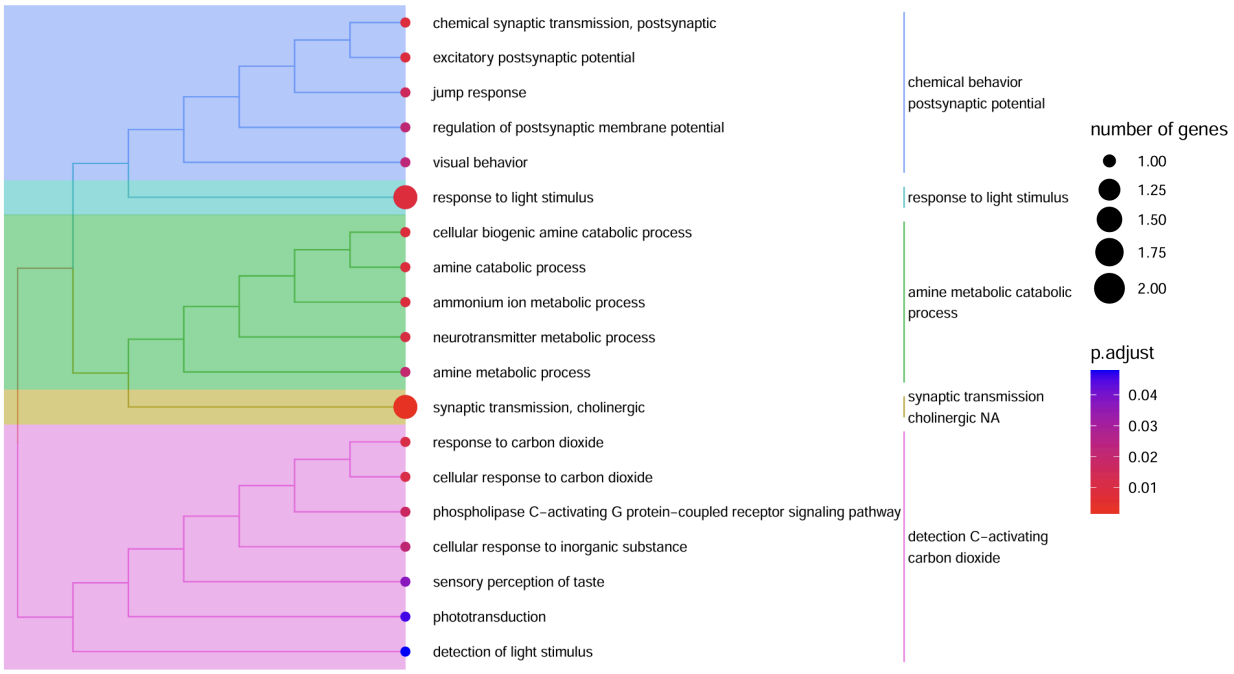
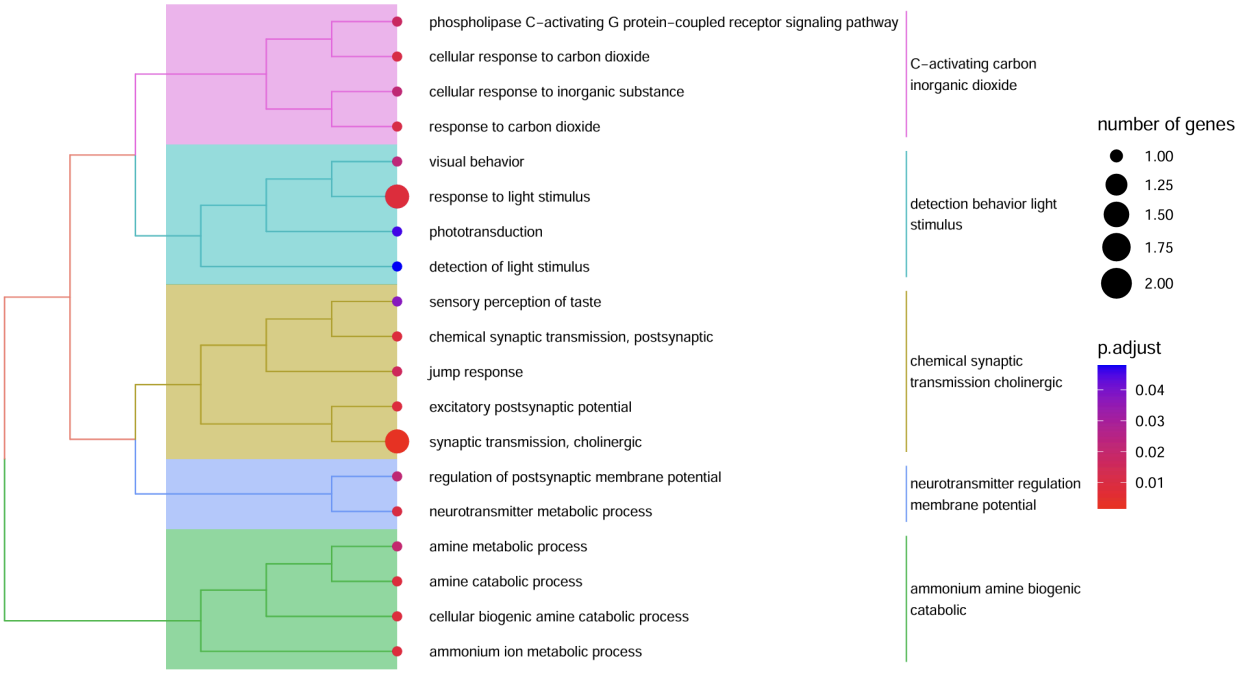


Figure 7: Hierarchical tree plot of synaptic AS GO annotation across all sub-ontologies for adult males vs. females using JC similarity. Overall parental pathway names and sub-clusters indicated by color blocks. Node size reflects the number of unique differential synaptic AS isoforms, while node color indicates the adjusted p-value. The top 30 GO terms (by significance) across BP, CC, and MF sub-ontologies are shown, clustered using Ward D2 algorithm.

A


ADULT MvF SYNAPTIC AS BP PATHWAY: JC METHOD
chemical synaptic transmission, postsynaptic
excitatory postsynaptic potential
jump response
regulation of postsynaptic membrane potential
visual behavior
response to light stimulus
cellular biogenic amine catabolic process
amine catabolic process
ammonium ion metabolic process
neurotransmitter metabolic process
amine metabolic process
synaptic transmission, cholinergic
response to carbon dioxide
cellular response to carbon dioxide
phospholipase C−activating G protein−coupled receptor signaling pathway
cellular response to inorganic substance
sensory perception of taste
phototransduction
detection of light stimulus
chemical behavior postsynaptic potential
response to light stimulus
amine metabolic catabolic process
synaptic transmission cholinergic NA
detection C−activating carbon dioxide
number of genes
1.00
1.25
1.50
1.75
2.00
p.adjust
0.04
0.03
0.02
0.01

B


ADULT MvF SYNAPTIC AS BP PATHWAY: RESNIK METHOD
phospholipase C−activating G protein−coupled receptor signaling pathway
cellular response to carbon dioxide
cellular response to inorganic substance
response to carbon dioxide
visual behavior
response to light stimulus
phototransduction
detection of light stimulus
sensory perception of taste
chemical synaptic transmission, postsynaptic
jump response
excitatory postsynaptic potential
synaptic transmission, cholinergic
regulation of postsynaptic membrane potential
neurotransmitter metabolic process
amine metabolic process
amine catabolic process
cellular biogenic amine catabolic process
ammonium ion metabolic process
C−activating carbon inorganic dioxide
detection behavior light stimulus
chemical synaptic transmission cholinergic
neurotransmitter regulation membrane potential
ammonium amine biogenic catabolic
number of genes
1.00
1.25
1.50
1.75
2.00
p.adjust
0.04
0.03
0.02
0.01

C

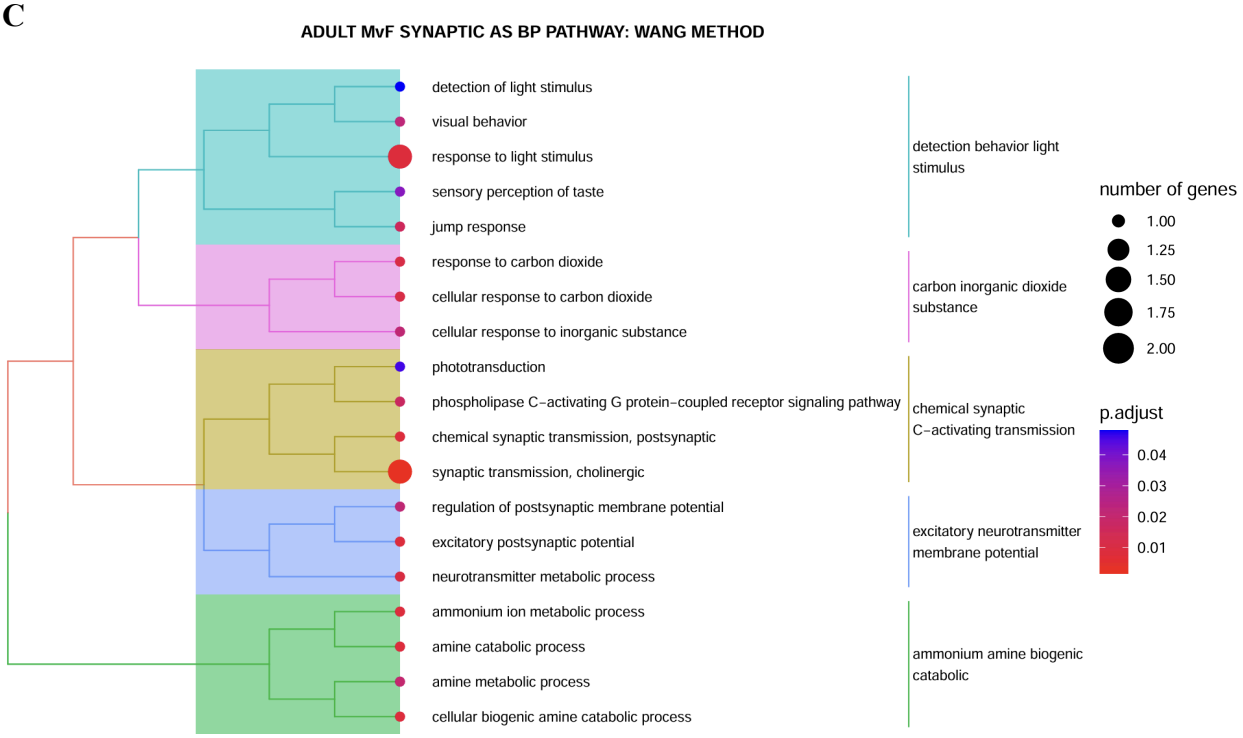
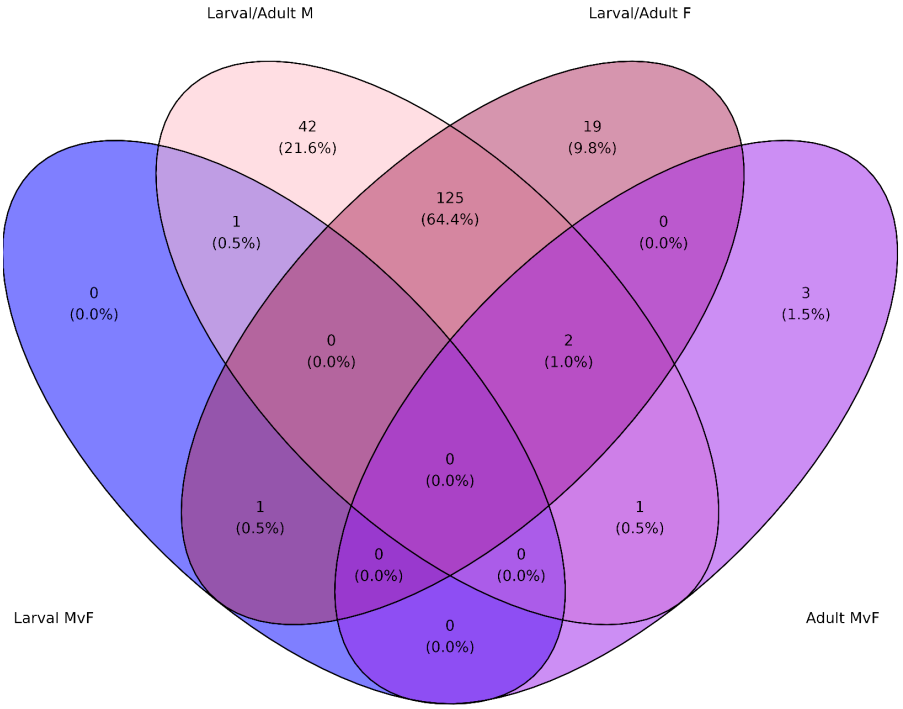


Figure 8: Hierarchical tree plot of synaptic AS GO BP sub-ontology terms for adult males vs. females. (A) Tree plot using JC method. **(B)** Tree plot using the Resnik method. **(C)** Tree plot using Wang method. Overall parental pathway names and sub-clusters indicated by color blocks. Node size reflects the number of unique differential synaptic AS isoforms, while node color indicates adjusted p-value. All significant GO terms in BP sub-ontology are shown, hierarchically clustered using Ward D2 algorithm.

A



B

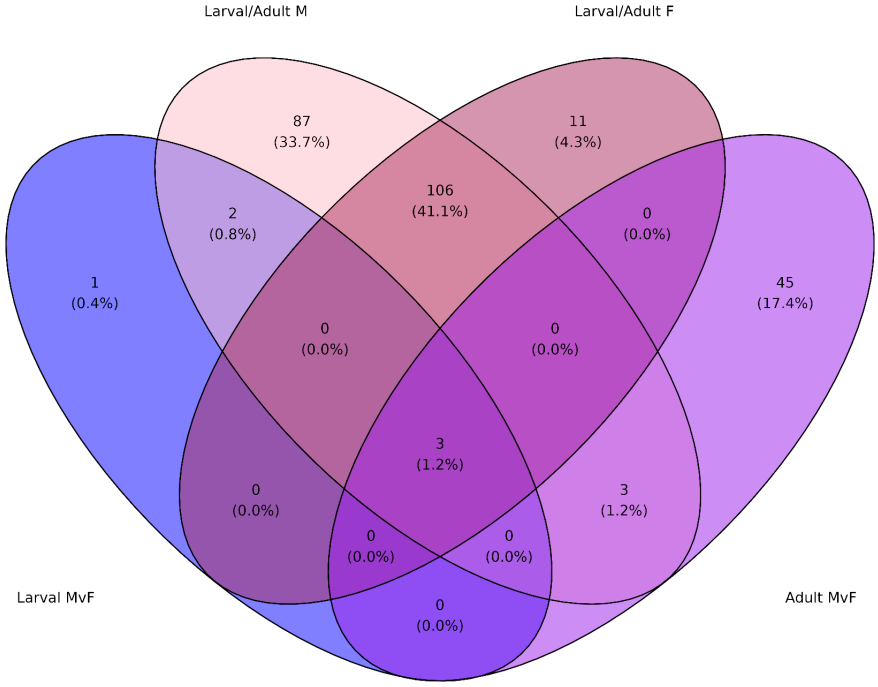
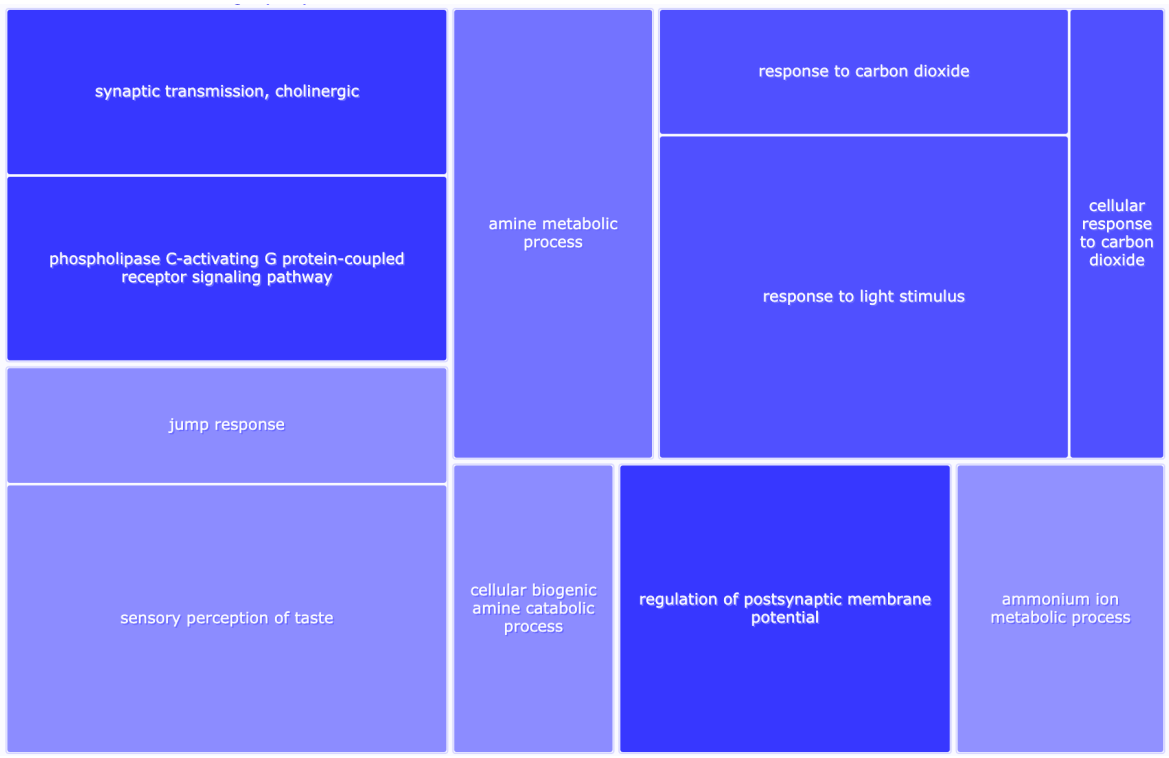


Figure 9: Shared and unique synaptic transcripts and GO terms across 4 comparisons. (A) Venn diagram of unique and shared synaptic transcript isoforms between four differential splicing comparisons. **(B)** Venn diagram of unique and shared enriched GO terms of corresponding transcripts between four differential splicing comparisons.

A


synaptic transmission, cholinergic
phospholipase C-activating G protein-coupled receptor signaling pathway
jump response
sensory perception of taste
amine metabolic process
response to carbon dioxide
response to light stimulus
cellular response to carbon dioxide
cellular biogenic amine catabolic process
regulation of postsynaptic membrane potential
ammonium ion metabolic process

B


rhabdomere
neuron spine
asymmetric synapse
extrinsic component of membrane
extrinsic component of plasma membrane
plasma membrane signaling receptor complex
receptor complex
GTPase complex

C

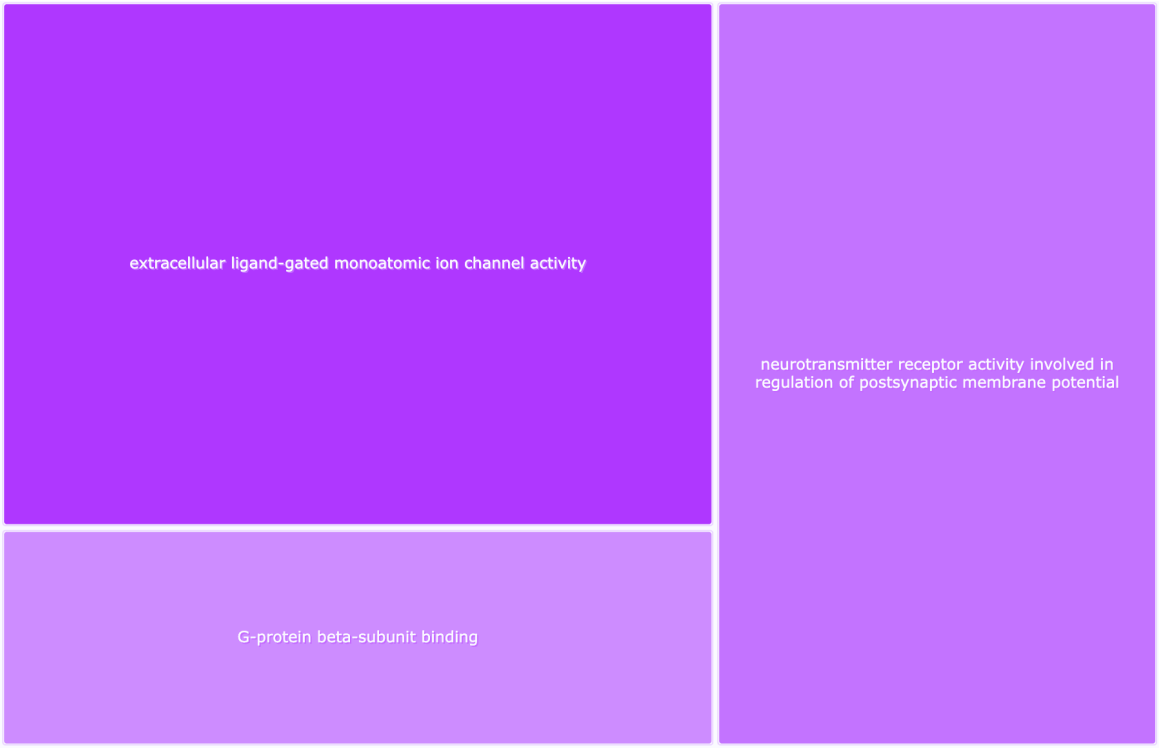


Figure 10: Tree plots of unique GO terms across each sub-ontology for adult male vs. female comparison. (A) BP-enriched GO terms. **(B)** CC-enriched GO terms. **(C)** MF-enriched GO terms. Intensity of color represents terms belonging to the same sub-pathway. The area of the rectangle represents the amount of enrichment, with larger areas representing greater relative enrichment between terms.

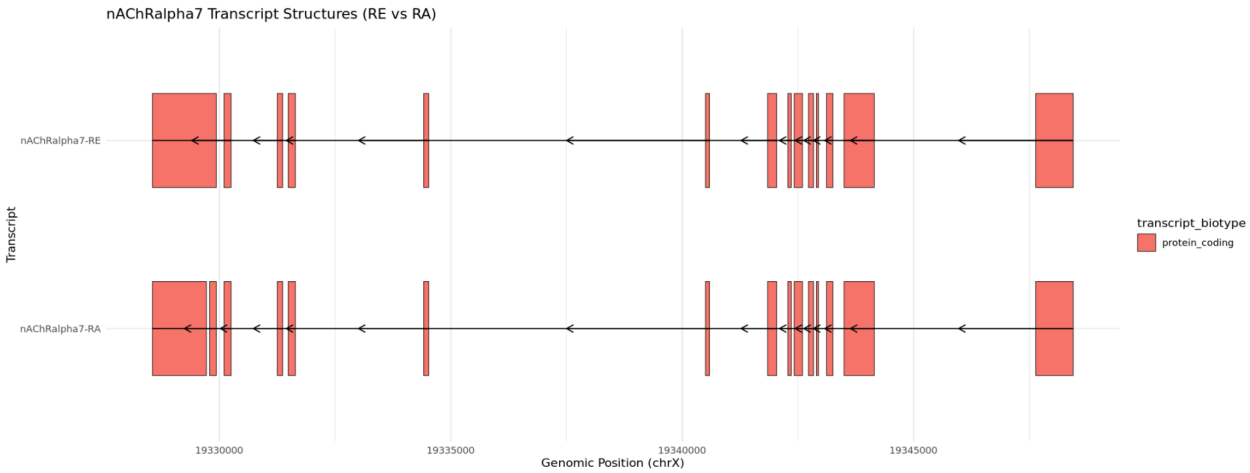


Figure 11: Structure of nAChRalpha7-RE and nAChRalpha7-RA transcript isoforms. Exonic regions are red rectangles, while intronic regions are represented by black lines. Genomic position of the transcripts is denoted on the x-axis.

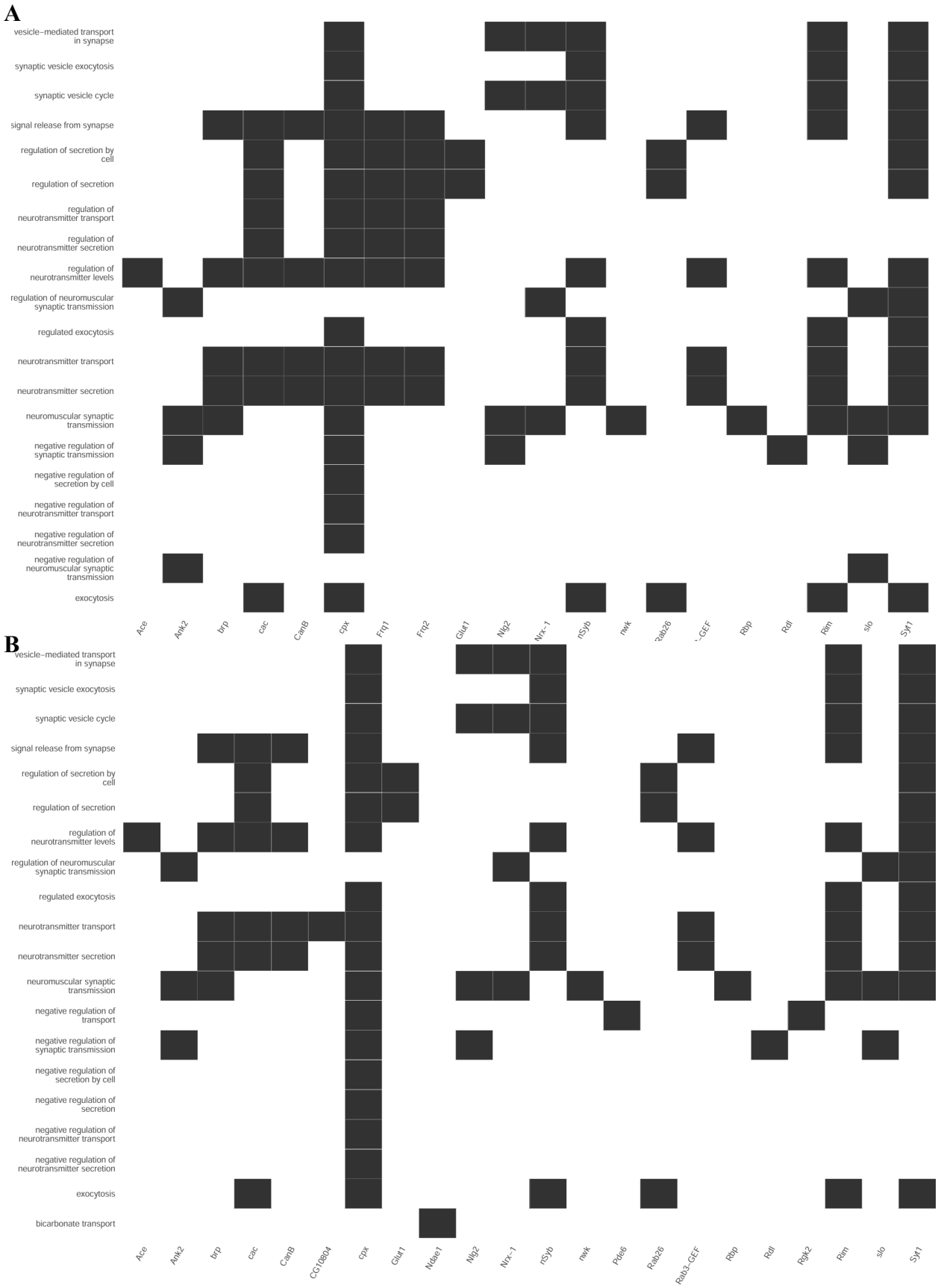

A
vesicle−mediated transport in synapse
synaptic vesicle exocytosis
synaptic vesicle cycle
signal release from synapse
regulation of secretion by cell
regulation of secretion
regulation of neurotransmitter transport
regulation of neurotransmitter secretion
regulation of neurotransmitter levels
regulation of neuromuscular synaptic transmission
regulated exocytosis
neurotransmitter transport
neurotransmitter secretion
neuromuscular synaptic transmission
negative regulation of synaptic transmission
negative regulation of secretion by cell
negative regulation of neurotransmitter transport
negative regulation of neurotransmitter secretion
negative regulation of neuromuscular synaptic transmission
exocytosis
Ace
Ank2
brp
cac
CamB
cpx
Frq1
Frq2
Glut1
Nlg2
Nrx−1
nSyb
nwk
Rab26
Rbp
Rdl
Rim
slo
Syt1
B
vesicle−mediated transport in synapse
synaptic vesicle exocytosis
synaptic vesicle cycle
signal release from synapse
regulation of secretion by cell
regulation of secretion
regulation of neurotransmitter levels
regulation of neuromuscular synaptic transmission
regulated exocytosis
neurotransmitter transport
neurotransmitter secretion
neuromuscular synaptic transmission
negative regulation of transport
negative regulation of synaptic transmission
negative regulation of secretion by cell
negative regulation of secretion
negative regulation of neurotransmitter transport
negative regulation of neurotransmitter secretion
exocytosis
bicarbonate transport
Ace
Ank2
brp
cac
CamB
CG10804
cpx
Glut1
Ndae1
Nlg2
Nrx−1
nSyb
nwk
Pde6
Rab26
Rab3−GEF
Rbp
Rdl
Rgk2
Rim
slo
Syt1

Figure S1: GO pathways of synaptic isoforms are differentially enriched within each sex. Heat plots of differentially enriched GO pathways within each sex, with parental genes of differentially spliced synaptic transcript on x-axis and GO annotations labeled on the y-axis. **(A)** Third instar vs. adult male comparison. **(B)** Third instar vs. adult female comparison. *cpx*, *nSyb*, *Rim* and *Syt1* represented across the most enrichment pathways for both sexes (p-adj. < 0.05).

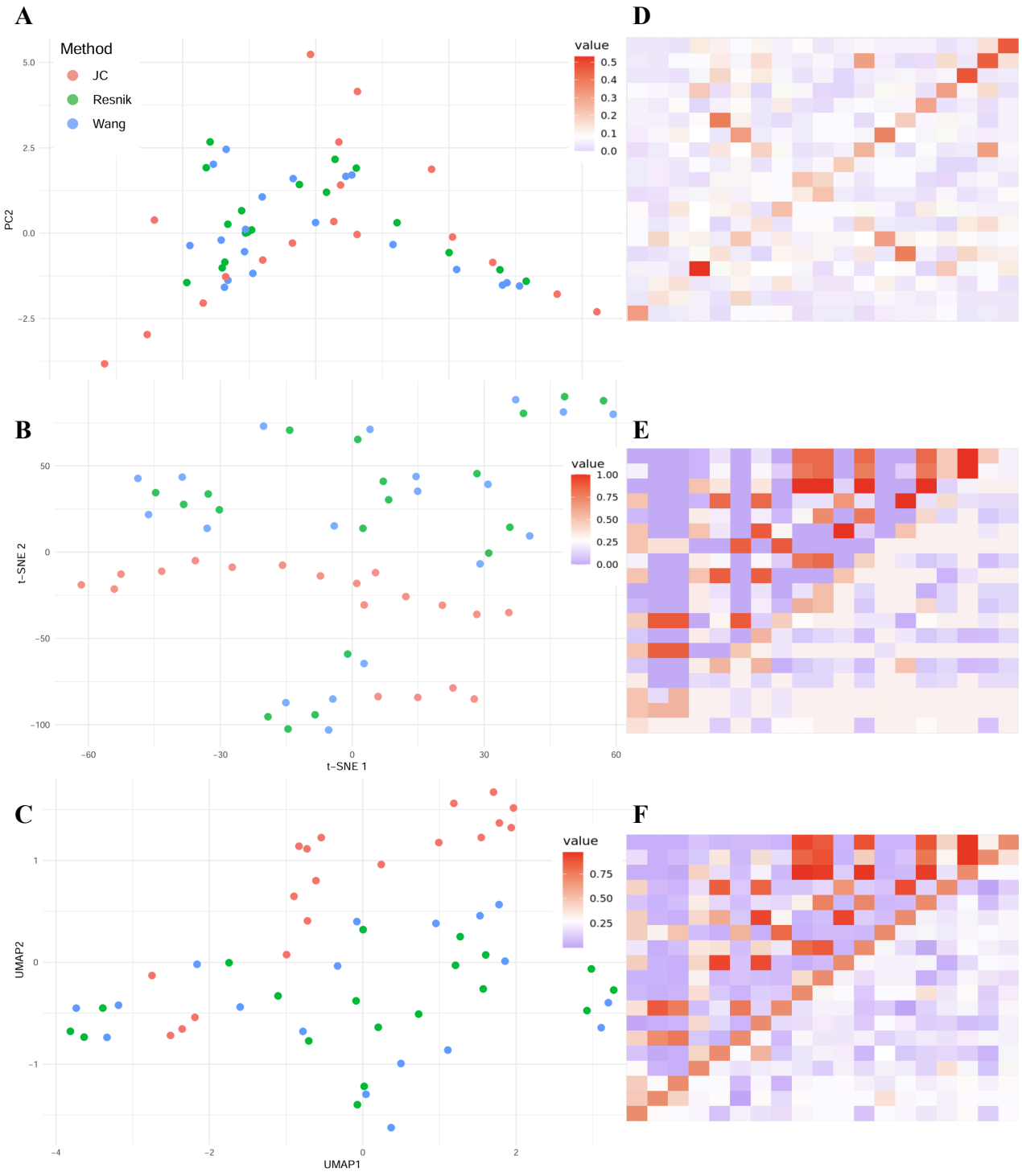


Figure S2: Resnik and Wang similarity methods cluster together in low-dimensional space. Dimensionality reduction visualizations of BP sub-ontology comparing JC, Resnik and Wang methods. **(A-C)** First two components using **(A)** principal component analysis (PCA), **(B)** t-distributed stochastic neighbor embedding (t-SNE) and **(C)** uniform manifold approximation projection (UMAP). **(D-F)** Heat plots of enriched BP pathway similarity matrices comparing **(D)** Resnik-Wang methods, **(E)** Resnik-JC methods and **(F)** Wang-JC methods.

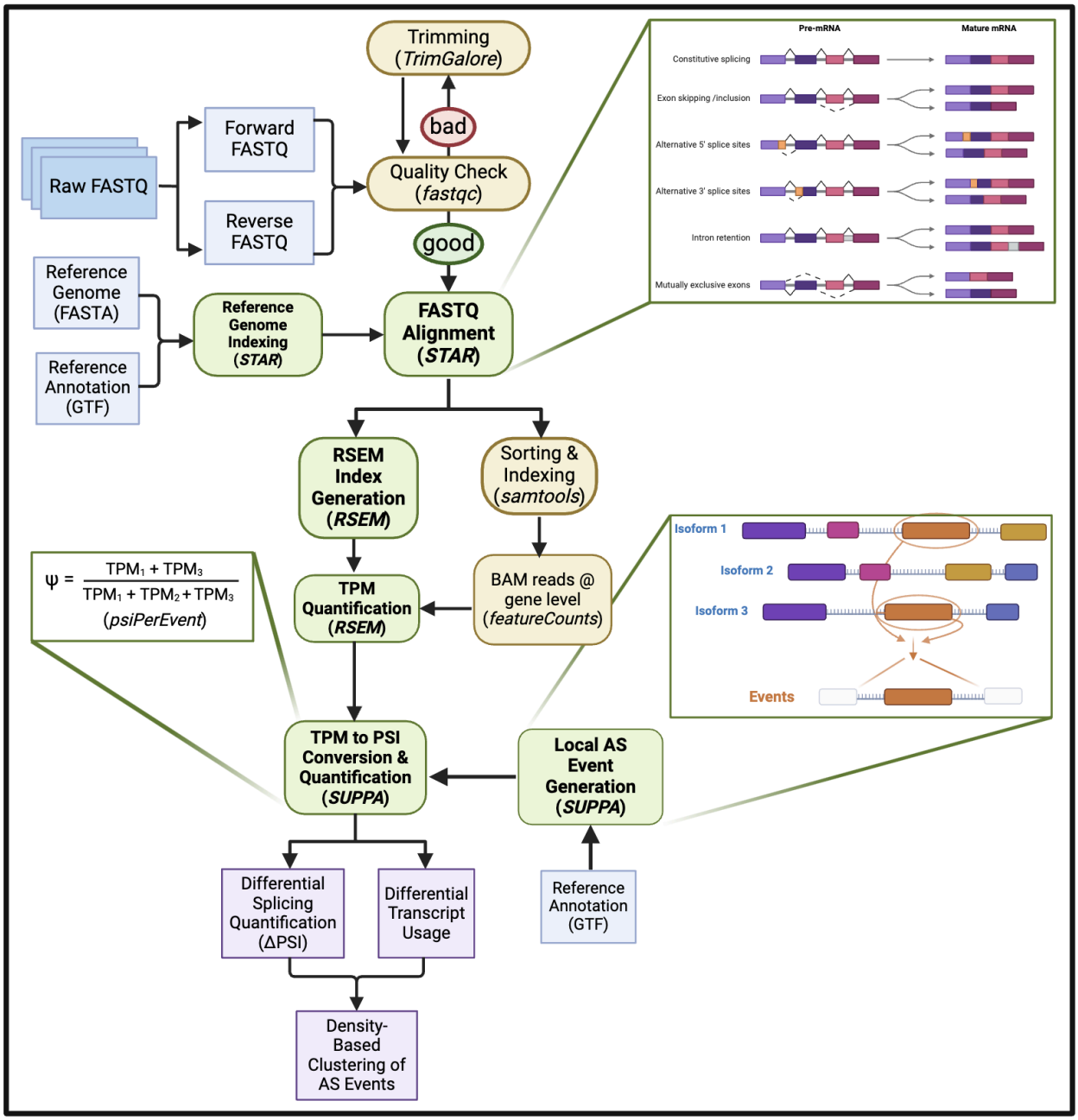


Figure S3: Bulk RNA-seq data processing and downstream analysis pipeline. Raw FASTQ files were first evaluated using QC metrics and aligned to a STAR-indexed reference genome. Splicing events were quantified in TPM using sorted BAM reads, converted to PSI, and then differential splicing was computed using SUPPA. Created with Biorender.com.

2.6 Tables

Gene Symbol	Event	Significance	Genomic Coordinates
Octalpha2R	RI	Male-Upregulated	3R:18799280:18806606-18806694:18807290:-
nAChRα7	RI	Female-Upregulated	X:19328554:19329720-19329790:19329933:-
nAChRα7	RI	Male-Upregulated	X:19347638:19347733-19347836:19348447:-
Ace	AL	Female-Upregulated	3R:13222951:13223378-13232679:13228242:13229770-13232679:-
dpr15	A3	Male-Upregulated	3R:12156621-12158534:12156605-12158534:-
Ggamma30A	A5	Female-Upregulated	2L:9256814-9256893:9256696-9256893:+

Table 1: Significant differential AS of synaptic genes between adult males and females.

GO Sub-ontology	Resnik-Wang (p)	Resnik-JC (p)	Wang-JC (p)
BP	**0.9131/0.8696**	0.2354/0.2654	0.1895/0.2517
CC	**0.4499/0.3461**	0.2286/0.1567	<0.01/0.0561
MF	0.3050/**0.3696**	**0.4271**/0.2075	<0.01/0.1787

Table 2: Pearson/Spearman correlations for semantic similarity matrices across GO sub-ontologies using JC, Resnik and Wang methods. The best results across all method comparisons are bolded and underlined. The best recorded individual metric for each sub-ontology is bolded.

2.7 References

Andrews, Simon. (2017) 2025. “S-Andrews/FastQC.” Java. https://github.com/s-andrews/FastQC.

Angarola, Brittany Lynn, and Olga Anczuków. 2021. “Splicing Alterations in Healthy Aging and Disease.” *Wiley Interdisciplinary Reviews. RNA* 12 (4): e1643. https://doi.org/10.1002/wrna.1643.

Ankerst, Mihael, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. 1999. “OPTICS: Ordering Points to Identify the Clustering Structure.” In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, 49–60. SIGMOD ’99. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/304182.304187.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. “Near-Optimal Probabilistic RNA-Seq Quantification.” *Nature Biotechnology* 34 (5): 525–27. https://doi.org/10.1038/nbt.3519.

Broide, R. S., and F. M. Leslie. 1999. “The Alpha7 Nicotinic Acetylcholine Receptor in Neuronal Plasticity.” *Molecular Neurobiology* 20 (1): 1–16. https://doi.org/10.1007/BF02741361.

Buza, Teresia J., Fiona M. McCarthy, Nan Wang, Susan M. Bridges, and Shane C. Burgess. 2008. “Gene Ontology Annotation Quality Analysis in Model Eukaryotes.” *Nucleic Acids Research* 36 (2): e12. https://doi.org/10.1093/nar/gkm1167.

Chen, Wenli, Weiwei Zhou, Qianqian Li, and Xiuguang Mao. 2023. “Sex Differences in Gene Expression and Alternative Splicing in the Chinese Horseshoe Bat.” *PeerJ* 11 (April):e15231. https://doi.org/10.7717/peerj.15231.

Cheng, Liang, Hengqiang Zhao, Pingping Wang, Wenyang Zhou, Meng Luo, Tianxin Li, Junwei Han, Shulin Liu, and Qinghua Jiang. 2019. “Computational Methods for Identifying Similar Diseases.” *Molecular Therapy. Nucleic Acids* 18 (September):590–604. https://doi.org/10.1016/j.omtn.2019.09.019.

Cieslak, Matthew C., Ann M. Castelfranco, Vittoria Roncalli, Petra H. Lenz, and Daniel K. Hartline. 2020. “T-Distributed Stochastic Neighbor Embedding (t-SNE): A Tool for Eco-Physiological Transcriptomic Analysis.” *Marine Genomics* 51 (June):100723. https://doi.org/10.1016/j.margen.2019.100723.

Costantini, E., C. Carrarini, P. Borrelli, M. De Rosa, D. Calisi, S. Consoli, D. D’Ardes, et al. 2023. “Different Peripheral Expression Patterns of the Nicotinic Acetylcholine Receptor in Dementia with Lewy Bodies and Alzheimer’s Disease.” *Immunity & Ageing* 20 (1): 3. https://doi.org/10.1186/s12979-023-00329-9.

Coughlin, Jennifer M., Yong Du, Hailey B. Rosenthal, Stephanie Slania, Soo Min Koo, Andrew Park, Ghedem Solomon, et al. 2018. “The Distribution of the Alpha7 Nicotinic Acetylcholine Receptor in Healthy Aging: An *in Vivo* Positron Emission Tomography Study with [18F]ASEM.” *NeuroImage* 165 (January):118–24. https://doi.org/10.1016/j.neuroimage.2017.10.009.

Danecek, Petr, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10 (2): giab008. https://doi.org/10.1093/gigascience/giab008.

Davis, Emily J., Caroline W. Solsberg, Charles C. White, Elena Miñones-Moyano, Marina Sirota, Lori Chibnik, David A. Bennett, Philip L. De Jager, Jennifer S. Yokoyama, and Dena B. Dubal. 2021. "Sex-Specific Association of the X Chromosome With Cognitive Change and Tau Pathology in Aging and Alzheimer Disease." *JAMA Neurology* 78 (10): 1–6. https://doi.org/10.1001/jamaneurol.2021.2806.

Deschênes, Mathieu, and Benoit Chabot. 2017. "The Emerging Role of Alternative Splicing in Senescence and Aging." *Aging Cell* 16 (5): 918–33. https://doi.org/10.1111/acel.12646.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics (Oxford, England)* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Durinck, Steffen, Paul T. Spellman, Ewan Birney, and Wolfgang Huber. 2009. "Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor Package biomaRt." *Nature Protocols* 4 (8): 1184–91. https://doi.org/10.1038/nprot.2009.97.

Fayuk, Dmitriy, and Jerrel L Yakel. 2007. "Dendritic Ca2+ Signalling Due to Activation of A7-Containing Nicotinic Acetylcholine Receptors in Rat Hippocampal Neurons." *The Journal of Physiology* 582 (Pt 2): 597–611. https://doi.org/10.1113/jphysiol.2007.135319.

Freedman, Robert, Michael Hall, Lawrence E. Adler, and Sherry Leonard. 1995. "Evidence in Postmortem Brain Tissue for Decreased Numbers of Hippocampal Nicotinic Receptors in Schizophrenia." *Biological Psychiatry* 38 (1): 22–33. https://doi.org/10.1016/0006-3223(94)00252-X.

Gan, Mingxin. 2014. "Correlating Information Contents of Gene Ontology Terms to Infer Semantic Similarity of Gene Products." *Computational and Mathematical Methods in Medicine* 2014:891842. https://doi.org/10.1155/2014/891842.

Guan, Z. Z., X. Zhang, R. Ravid, and A. Nordberg. 2000. “Decreased Protein Levels of Nicotinic Receptor Subunits in the Hippocampus and Temporal Cortex of Patients with Alzheimer’s Disease.” *Journal of Neurochemistry* 74 (1): 237–43. https://doi.org/10.1046/j.1471-4159.2000.0740237.x.

Harrison, Peter W, M Ridwan Amode, Olanrewaju Austine-Orimoloye, Andrey G Azov, Matthieu Barba, If Barnes, Arne Becker, et al. 2024. “Ensembl 2024.” *Nucleic Acids Research* 52 (D1): D891–99. https://doi.org/10.1093/nar/gkad1049.

Hervé, Abdi, and Lynne J. Williams. 2010. “Principal Component Analysis.” *WIREs Computational Statistics* 2 (4): 433–59. https://doi.org/10.1002/wics.101.

Izuo, Naotaka, Nobuhiro Watanabe, Yoshihiro Noda, Takashi Saito, Takaomi C. Saido, Koutaro Yokote, Harumi Hotta, and Takahiko Shimizu. 2023. “Insulin Resistance Induces Earlier Initiation of Cognitive Dysfunction Mediated by Cholinergic Deregulation in a Mouse Model of Alzheimer’s Disease.” *Aging Cell* 22 (11): e13994. https://doi.org/10.1111/acel.13994.

Jaccard, Paul. 1902. “Distribution comparée de la flore alpine dans quelques régions des Alpes occidentales et orientales.” *Bulletin de la Murithienne*, no. 31, 81–92.

Kang, Hyo Jung, Yuka Imamura Kawasawa, Feng Cheng, Ying Zhu, Xuming Xu, Mingfeng Li, André M. M. Sousa, et al. 2011. “Spatio-Temporal Transcriptome of the Human Brain.” *Nature* 478 (7370): 483–89. https://doi.org/10.1038/nature10523.

Kentro, James A., Gunjan Singh, Tuan M. Pham, Justin Currie, Saniya Khullar, Audrey T. Medeiros, Maria Tsiarli, Erica Larschan, and Kate M. O’Connor-Giles. 2025. “Conserved Transcription Factors Coordinate Synaptic Gene Expression through Repression.” *bioRxiv*, February, 2024.10.30.621128. https://doi.org/10.1101/2024.10.30.621128.

Krueger, Felix, Frankie James, Phil Ewels, Ebrahim Afyounian, Michael Weinstein, Benjamin Schuster-Boeckler, Gert Hulselmans, and sclamons. 2023. “FelixKrueger/TrimGalore: V0.6.10 - Add Default Decompression Path.” Zenodo. https://doi.org/10.5281/zenodo.7598955.

Legendre, P., and Louis Legendre. 2012. *Numerical Ecology*. Elsevier.

Li, Bo, and Colin N. Dewey. 2011. “RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome.” *BMC Bioinformatics* 12 (1): 323. https://doi.org/10.1186/1471-2105-12-323.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7): 923–30. https://doi.org/10.1093/bioinformatics/btt656.

Lopez-Lee, Chloe, Eileen Ruth S. Torres, Gillian Carling, and Li Gan. 2024. “Mechanisms of Sex Differences in Alzheimer’s Disease.” *Neuron* 112 (8): 1208–21. https://doi.org/10.1016/j.neuron.2024.01.024.

Maaten, L.J.P. van der, and G.E. Hinton. 2008. “Visualizing High-Dimensional Data Using t-SNE.” *Journal of Machine Learning Research* 9 (nov): 2579–2605.

McInnes, Leland, John Healy, and James Melville. 2020. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.” arXiv. https://doi.org/10.48550/arXiv.1802.03426.

Murtagh, Fionn, and Pierre Legendre. 2014. “Ward’s Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm.” *Journal of Classification* 31 (3): 274–95. https://doi.org/10.1007/s00357-014-9161-z.

Nacer, Samir A., Simone Otto, Ayland C. Letsinger, Jemma Strauss DeFilipp, Viktoriya D. Nikolova, Natallia V. Riddick, Korey D. Stevanovic, Jesse D. Cushman, and Jerrel L. Yakel.

2020. “Loss of A7 Nicotinic Acetylcholine Receptors in GABAergic Interneurons Causes Sex-Dependent Impairments in Postnatal Neurogenesis and Cognitive and Social Behavior.” bioRxiv. https://doi.org/10.1101/2020.03.16.994111.

Öztürk-Çolak, Arzu, Steven J Marygold, Giulia Antonazzo, Helen Attrill, Damien Goutte-Gattat, Victoria K Jenkins, Beverley B Matthews, et al. 2024. “FlyBase: Updates to the Drosophila Genes and Genomes Database.” *Genetics* 227 (1): iyad211. https://doi.org/10.1093/genetics/iyad211.

Pan, Qun, Ofer Shai, Leo J. Lee, Brendan J. Frey, and Benjamin J. Blencowe. 2008. “Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing.” *Nature Genetics* 40 (12): 1413–15. https://doi.org/10.1038/ng.259.

Pan, Yan, Zhu, Jing, Cong, Lina, Bai, Yang, Ma, Yaxin, and Yue and Yang. 2021. “Expression of nAChRα7 Receptor in Model Rats with Parkinson’s Disease Dementia.” *Biotechnology & Biotechnological Equipment* 35 (1): 117–23. https://doi.org/10.1080/13102818.2020.1829500.

“Finding Groups in Data: An Introduction To Cluster Analysis.” 2024. *ResearchGate*, October. https://doi.org/10.2307/2532178.

Pohanka, Miroslav. 2012. “Alpha7 Nicotinic Acetylcholine Receptor Is a Target in Pharmacology and Toxicology.” *International Journal of Molecular Sciences* 13 (2): 2219–38. https://doi.org/10.3390/ijms13022219.

Rabah, Yasmine, Jean-Paul Berwick, Nisrine Sagar, Laure Pasquer, Pierre-Yves Plaçais, and Thomas Preat. 2025. “Astrocyte-to-Neuron H2O2 Signalling Supports Long-Term Memory Formation in Drosophila and Is Impaired in an Alzheimer’s Disease Model.” *Nature Metabolism* 7 (2): 321–35. https://doi.org/10.1038/s42255-024-01189-3.

Rabah, Yasmine, Nisrine Sagar, Laure Pasquer, Pierre-Yves Plaçais, and Thomas Preat. 2023. "Long-Term Memory Formation Depends on an Astrocyte-to-Neuron H2O2 Signaling." bioRxiv. https://doi.org/10.1101/2023.07.11.548505.

Resnik, P. 1999. "Semantic Similarity in a Taxonomy: An Information-Based Measure and Its Application to Problems of Ambiguity in Natural Language." *Journal of Artificial Intelligence Research* 11 (July):95–130. https://doi.org/10.1613/jair.514.

Rosenthal, Justin S., Jun Yin, Jingce Lei, Anupama Sathyamurthy, Jacob Short, Caixia Long, Emma Spillman, Chengyu Sheng, and Quan Yuan. 2021. "Temporal Regulation of Nicotinic Acetylcholine Receptor Subunits Supports Central Cholinergic Synapse Development in Drosophila." *Proceedings of the National Academy of Sciences of the United States of America* 118 (23): e2004685118. https://doi.org/10.1073/pnas.2004685118.

Scotti, Marina M., and Maurice S. Swanson. 2016. "RNA Mis-Splicing in Disease." *Nature Reviews Genetics* 17 (1): 19–32. https://doi.org/10.1038/nrg.2015.3.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PloS One* 6 (7): e21800. https://doi.org/10.1371/journal.pone.0021800.

Trincado, Juan L., Juan C. Entizne, Gerald Hysenaj, Babita Singh, Miha Skalic, David J. Elliott, and Eduardo Eyras. 2018. "SUPPA2: Fast, Accurate, and Uncertainty-Aware Differential Splicing Analysis across Multiple Conditions." *Genome Biology* 19 (1): 40. https://doi.org/10.1186/s13059-018-1417-1.

Villa, Alessandro, Paolo Gelosa, Laura Castiglioni, Mauro Cimino, Nicoletta Rizzi, Giovanna Pepe, Federica Lolli, et al. 2018. "Sex-Specific Features of Microglia from Adult Mice." *Cell Reports* 23 (12): 3501–11. https://doi.org/10.1016/j.celrep.2018.05.048.

Wallace, Tanya L., and Richard H. P. Porter. 2011. “Targeting the Nicotinic Alpha7 Acetylcholine Receptor to Enhance Cognition in Disease.” *Biochemical Pharmacology*, Nicotinic Acetylcholine Receptors as Therapeutic Targets: Emerging Frontiers in Basic Research and Clinical Science (Satellite to the 2011 Meeting of the Society for Neuroscience), 82 (8): 891–903. https://doi.org/10.1016/j.bcp.2011.06.034.

Wang, Eric T., Rickard Sandberg, Shujun Luo, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore, Gary P. Schroth, and Christopher B. Burge. 2008. “Alternative Isoform Regulation in Human Tissue Transcriptomes.” *Nature* 456 (7221): 470–76. https://doi.org/10.1038/nature07509.

Wang, James Z., Zhidian Du, Rapeeporn Payattakool, Philip S. Yu, and Chin-Fu Chen. 2007. “A New Method to Measure the Semantic Similarity of GO Terms.” *Bioinformatics* 23 (10): 1274–81. https://doi.org/10.1093/bioinformatics/btm087.

Ward Jr., Joe H. 1963. “Hierarchical Grouping to Optimize an Objective Function.” *Journal of the American Statistical Association* 58 (301): 236–44. https://doi.org/10.1080/01621459.1963.10500845.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, Davis Vaughan, Posit Software, and PBC. 2023. “Dplyr: A Grammar of Data Manipulation.” https://cran.r-project.org/web/packages/dplyr/index.html.

Wu, Tianzhi, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, Tingze Feng, et al. 2021. “clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data.” *The Innovation* 2 (3). https://doi.org/10.1016/j.xinn.2021.100141.

Xu, Shuangbin, Erqiang Hu, Yantong Cai, Zijing Xie, Xiao Luo, Li Zhan, Wenli Tang, et al. 2024. “Using clusterProfiler to Characterize Multiomics Data.” *Nature Protocols* 19 (11): 3292–3320. https://doi.org/10.1038/s41596-024-01020-z.

Xue, Xiaoli, Wei Zhang, and Anjing Fan. 2023. “Comparative Analysis of Gene Ontology-Based Semantic Similarity Measurements for the Application of Identifying Essential Proteins.” *PLOS ONE* 18 (4): e0284274. https://doi.org/10.1371/journal.pone.0284274.

Yang, Yang, Hongjian Sun, Yu Zhang, Tiefu Zhang, Jialei Gong, Yunbo Wei, Yong-Gang Duan, et al. 2021. “Dimensionality Reduction by UMAP Reinforces Sample Heterogeneity Analysis in Bulk Transcriptomic Data.” *Cell Reports* 36 (4): 109442. https://doi.org/10.1016/j.celrep.2021.109442.

Yu, Guangchuang. 2020. “Gene Ontology Semantic Similarity Analysis Using GOSemSim.” In *Stem Cell Transcriptional Networks: Methods and Protocols*, edited by Benjamin L. Kidder, 207–15. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-0301-7_11.

Chapter 3: The evolutionary conservation of temporally-coordinated synaptic gene expression in human cells

3.1 Introduction

In Chapter 2, I identified sex and developmentally-biased synaptic AS in *Drosophila melanogaster* through bulk transcriptomics analysis. These findings raised the question of whether the regulatory dynamics displayed for synaptic genes are conserved in more complex species. In this chapter, I explore the evolutionary conservation of temporally-coordinated synaptic gene expression in human cells using single-cell resolution multi-omics data.

The development of the CNS in both invertebrate and mammalian species is a complex and involved process. In *Drosophila*, larvae undergo widespread synaptic restructuring during pupation, resulting in a reconstructed adult nervous system. While well-documented, the regulatory mechanisms underlying this spatiotemporally-coordinated development have largely remained elusive. Recent work by Kentro et al. identified the chromatin regulators of DEAF1 and CLAMP as key repressors of temporally-regulated synaptogenesis and synaptic gene expression across *Drosophila* development (Kentro et al., 2024). Strikingly, this synaptic gene regulation appears in mouse embryogenesis, with demonstrated temporal coordination of both pan-neuronal and neuron-subtype-specific synaptic gene transcription (Cardoso-Moreira et al., 2019)—suggesting that the developmental mechanism of coordinated synaptic repression may be evolutionarily conserved (Figure 12, Kentro et al., 2024).

In humans, cortical synaptic density undergoes a rapid increase during the early postnatal period, followed by synaptic pruning and the stabilization of synaptic connections in adulthood (Huttenlocher et al., 1982; Bruer et al., 1999; Goyal and Raichle, 2013). Over the course of aberrant aging, like in the case of AD and other dementias, cortical synaptic density and brain volume rapidly decrease (Davies et al., 1987; Clare et al., 2010), possibly reflecting a loss of

synaptic gene expression. By characterizing synaptic gene regulation and chromatin accessibility at the single-cell resolution in human cells, I sought to determine whether the temporally coordinated synaptic gene expression and patterns of chromatin accessibility in *Drosophila* and mice were evolutionarily conserved across cerebral organoid development. In doing so, I provide further insight into the molecular pathways implicated in both healthy and aberrant synaptic development.

3.2 Results

3.2.1 Global synaptic gene ortholog expression is temporally coordinated across various stages of human cerebral organoid development

I took the 485 coordinately regulated synaptic genes identified by Kentro et al. (2024) and mapped them to orthologous human genes using the DRSC Integrative Ortholog Prediction Tool (Hu et al., 2021). After selecting for high and mid-matching orthologs, I was left with 332 orthologous human synaptic genes. To provide a robust characterization of synaptic gene expression, I initially processed four scRNA-seq cerebral organoid datasets of different cell lines, recorded time points, and with varying cell counts per condition obtained roughly 640,000 total cells for analysis. The datasets measured gene expression in organoids from day 22 to day 101 (Table 3, Sebastian et al. 2023), embryonic (EB) stage to 8 months (Table 4, Zenk et al., 2024), 23 days to 6 months (Table 5, Uzquiano et al., 2022), or 4 days to 61 days (Table 6, Fleck et al., 2022), respectively. After scRNA-seq data preprocessing, integration, and conversion to Seurat objects, I generated feature plots and dotplots for each dataset.

The generated feature plot for days 22 to 101 (Sebastian et al., 2023) revealed spatially and temporally distinct clusters of global synaptic gene expression (Figure 13A). In line with this observation, a dotplot of the global synaptic gene expression profile at each timepoint revealed

greater synaptic gene expression as the organoid developed (Figure 13B). Separating the day 101 time point into its 2 component replicates demonstrated low global gene expression in one replicate and the highest recorded global expression in the other, suggesting a potential issue with the read quality of the sparsely expressed replicate. In line with this observation, removing the replicate at each time point with the lowest level of synaptic expression showed an increasing pattern as the organoid developed. I next checked whether this pattern of increase in synaptic gene expression across development manifested in other datasets, starting with that of Zenk et al. (2024).

While interesting on its own, it was possible that the increasing global synaptic gene expression observed in the Sebastian et al. (2023) dataset stemmed from technical rather than biological factors. To account for the possible impact of batch effects, I used a dataset (Zenk et al., 2024) that recorded over a longer period of time and with greater frequency—spanning from pluripotent embryoid bodies recorded at day 5 (EB) to 8 months across 8 recorded timepoints. I observed spatially distinct clusters at each developmental stage, with increased global synaptic gene expression as the organoid approached the last recorded developmental timepoint (Figure 14). While not as spatially separated as in the first dataset, a dot plot of the recorded time points revealed increasing global synaptic gene expression and overall percentage of expressed synaptic genes (Figure 14B).

By testing a dataset with greater longitudinal breadth, I observed increasing synaptic gene expression in a developmental stage-specific manner. To confirm this pattern of synaptic gene expression across cell lines with similarly recorded timepoints, I analyzed a more expansive dataset (Uzquiano et al., 2022), measuring 8 time points from day 23 to 6 months. Unlike the prior tested datasets, the Uzquiano et al. (2022) data contained roughly 550,000 recorded cells across timepoints, suggesting that it may offer a more robust characterization of synaptic gene

expression at the cellular resolution than those tested prior. Due to computational limitations, I processed and then randomly downsampled the data, taking 10% of cells at each time point for downstream analysis. Surprisingly, the observed pattern of synaptic gene expression differed greatly from previously tested datasets, with no overt spatially segregated cell groups or increasing trend in synaptic gene expression as the organoid developed (Figure 15A). Similarly, later time points displayed a lower percentage of total synaptic genes expressed and lower average expression (Figure 15B). It is possible that the synaptic gene expression pattern observed resulted from the inability to fully preprocess roughly 550,000 cells—indicative of batch-related error, rather than reflective of all cellular synaptic expression profiles.

Despite the greater temporal breadth and cell count of the above datasets, I still wished to develop a greater understanding of the synaptic gene expression changes that occur *earlier* in cerebral organoid development. To do so, I analyzed scRNA-seq data from Zenk et al. (2024) that sampled 11 timepoints from day 4 to 61. As in the case of the Uzquiano et al. (2022) dataset, I observed spatially clustered synaptic gene expression in single cells, with increasing expression as the organoid developed (Figure 16A). The dotplot of global synaptic gene expression similarly conveyed this trend, with greater average expression and percentage of expressed synaptic genes as the organoid approached the last recorded time point (Figure 16B). Interestingly, I found that the top 10 most highly expressed synaptic genes varied greatly across developmental stages: *STOM*, *ARFGEF3*, *CDH1*, *SYT6,* and *HS3ST5* were highly expressed only at early developmental stages, while *SLC2A3*, *SCG5*, *EML4*, *APLP2,* and *NLGN4X* tended to be expressed throughout development (Figure 16C). The temporal clustering of these genes may indicate developmental stage-specific regulatory mechanisms, which impact gene expression during later stages.

3.2.2 Chromatin accessibility at synaptic gene loci in human cells oppose *Drosophila* developmental patterns

On its own, the scRNA-seq data revealed distinct patterns of synaptic gene expression across several different cerebral organoid lines. While insightful, the integration of scATAC-seq may provide more insight into the changes occurring to the 3D chromatin architecture of synaptic genes. Therefore, I took the paired scATAC-seq data from Fleck et al. (2022) to identify whether chromatin accessibility preceded increased synaptic gene expression at older developmental periods, as was observed in *Drosophila* development (Kentro et al. 2024). The feature plot and dotplot of the Fleck et al. (2022) dataset are shown in Figure 17. Unexpectedly, chromatin accessibility peaked at early time points, followed by a decrease in accessibility, and then a final increase at the last recorded time point. (Figure 17A-B).

To gain a better understanding of when changes to chromatin accessibility occur in relation to gene expression, I plotted the normalized module scores across developmental time points for both synaptic and random genes for the scRNA-seq and scATAC-seq data (Figure 17C). As expected, the pattern of synaptic chromatin accessibility and gene expression appeared distinct from the random genes set. However, unlike the temporal pattern of chromatin accessibility opening before synaptic gene expression that Kentro et al. (2024) observed, I found chromatin accessibility decreased as coordinated synaptic gene expression increased. This pattern of chromatin accessibility was directly opposed to *Drosophila* development, but the global increase in synaptic gene expression at later developmental time points appears across organoid lines, flies, and mice.

3.2.3 Sex-Biased Differentially Spliced Synaptic Isoforms in *Drosophila* Drive Glutamatergic Neuron Developmental Transitions in Human Cortex Tissue

To complement the bulk tissue and single-cell module score analyses described in the prior sections, I next examined the gene expression patterns of a few select synaptic genes using data from human brain tissue. I integrated the scRNA-seq and scATAC-seq data of human cerebral cortex tissue samples from gestational week 18 to 39 year old donors (Table 7), using an unsupervised deep learning method called scMultiNODE (Zhang et al., 2024). scMultiNODE is able to infer cell fate trajectory at the single-cell level using principles of optimal transport (OT). In OT, a model assumes that the gene expression (or chromatin accessibility) per cell at a given time point follows a probabilistic distribution, such that the origin and terminal fate of the cell are inferrable (Schiebinger et al., 2019). Coupling between sequential time points was calculated for each cell, and then extended across all timepoints by composing OT maps across the entirety of a time interval (Figure 18). After training scMultiNODE on the dataset, I constructed the least action path (LAP) between any two cell states, while minimizing action and transition time (Qiu et al., 2012; Wang et al., 2014; Qiu et al., 2022). I generated the LAP paths from individual cell starting points (t = 0) to developed oligodendrocyte and glutamatergic neuron populations (Figure 19A). After capturing all significant differentially expressed genes (DEGs) along each lineage trajectory, I found *STON2* and *KCNC2* to be the highest ranking human synaptic gene orthologs in oligodendrocytes and glutamatergic neurons, respectively. I selected these genes for further downstream analysis, in addition to *CHRNA7*, which is discussed below.

Returning briefly to Chapter 2, I observed sex-biased differential *nAChRalpha7* splicing variants unique to the adult fly developmental stage through bulk RNA-seq data analysis. Its human ortholog, *CHRNA7*, has been implicated in several CNS and neurodevelopmental disorders (Freedman et al., 1995; Court et al., 1999; Stephens et al., 2009; Ancín et al., 2011).

Due to these observations, it is plausible that *CHRNA7* (or *nAChRalpha7*) may have a potential role in sexually dimorphic neuronal maintenance and function. Given the involvement of *CHRNA7* in excitatory neurotransmission and mammalian neural development, I hypothesized that its synaptic gene expression trajectory in human cells may mirror key transitions in excitatory neuronal lineage.

To explore this hypothesis, I plotted the gene expression z-score values of *KCNC2* & *CHRNA7* for glutamatergic neuronal trajectory and *STON2* & *CHRNA7* for oligodendrocytes across developmental time points (Figure 19). In comparison to random genes, *STON2* displayed greater z-score variability across oligodendrocyte development, while *CHRNA7* only differed slightly from random genes (Figure 19B). On the other hand, both *KCNC2* and *CHRNA7* displayed much greater z-score variability across glutamatergic neuron development (Figure 19C). *CHRNA7* gene expression appeared to peak at the latest developmental timepoints in glutamatergic neuronal development, suggesting its potential role in adult-specific neural maintenance. The discrepancy in *CHRNA7* gene expression patterns between cell types also demonstrated a neuronal-enriched role in development. Together, these results reinforce the conserved role of *CHRNA7* and its fly ortholog *nAChRalpha7* in neuronal development and synaptic function, potentially in a temporal and sex-specific manner.

3.3 Discussion

In this chapter, I profiled the synaptic gene expression of 640,000 cells across five human cell-derived datasets. The progression of synaptic gene expression across developmental time points in human cerebral organoids generally reflected a conserved pattern of neural maturation *in vitro*. The coordinated increase in synaptic gene module scores across developmental time points in human cerebral organoids suggests that temporally coordinated synaptic gene

expression is a conserved hallmark of neurodevelopment. This result echoed the findings observed in both *Drosophila* and mouse models (Kentro et al., 2024), supporting a potential evolutionarily conserved role for temporally coordinating synaptic gene expression—even in the absence of fully developed brain architecture. The fact that this pattern of gene expression is recapitulated across multiple datasets and experimental designs—albeit with variation in replicate quality—supports the robustness of developmentally coordinated synaptic gene expression.

Interestingly, chromatin accessibility at the synaptic loci in Fleck et al. (2022) failed to mirror transcriptomic trends, diverging from the coordinated accessibility patterns observed in *Drosophila* development. This difference points to more complex regulatory layers than those present in either *Drosophila* or mouse—although transcriptional activation is temporally coordinated, upstream chromatin remodeling may occur earlier in human cell development—thereby highlighting a potential decoupling of epigenetic and transcriptional machinery in human organoid models. Alternatively, the process of generating cerebral organoids themselves, rather than analyzing tissue from autonomous organisms, may contribute to the observed differences in chromatin accessibility.

3.4 Materials and Methods

3.4.1 Human cerebral organoid temporal profile analysis

Data processing and quality control

To generate an analogous set of synaptic genes, I identified orthologs of the 485 *Drosophila* synaptic expression profile genes (Kentro et al., 2024) ranked as high or moderate according to the DRSC Integrative Ortholog Prediction Tool (IOPT) (Hu et al., 2021), finding 332 mid- and

high-matching orthologous human genes (Appendix A). I used these orthologs to generate an organoid synaptic expression profile across four separate scRNA-seq datasets. For later downstream analysis of the paired scRNA-seq and scATAC-seq data (Fleck et al., 2022), I additionally identified mid- and high-matching human orthologs of 485 random *Drosophila* genes (Kentro et al. 2024) using the DRSC IOPT, taking the top 332 human genes (Appendix B).

Human cerebral organoid scRNA-seq (Uzquiano et al., 2022; Sebastian et al., 2023; Zenk et al., 2024) and paired scRNA-seq & scATAC-seq data (Fleck et al., 2022) were processed with Seurat v5.0 (Hao et al., 2021) and Signac v1.1.0 (Stuart et al., 2021) to create Seurat objects. Using Seurat, the four cerebral organoid datasets were loaded as count matrices. Seurat objects were created and filtered with min.cells = 3 and min.features = 200, then merged. Prior to QC evaluation and in order to run downstream analyses, virtual memory was set to 16GB using options(future.globals.maxSize = 16 * 1024^3). The percentage of all-gene UMI mapping to mitochondrial (MT)-genes was used as a QC metric. Each merged Seurat object was normalized using SCTransform (Hafemeister & Satija, 2019), regressing on my mitochondrial gene mapping percentage. I used the 3000 most variable genes and then performed principal component analysis (PCA) using the RunPCA() function in Seurat.

Clustering and visualization

Using 30 PCs, I identified nearest neighbors with the Seurat function FindNeighbors(seurat_obj, dims = 1:30, reduction = "pca"), followed by using a shared nearest neighbor (SNN) (Levine et al., 2015; Xu and Su, 2015) modularity optimization-based clustering algorithm. For clustering, I manually checked the effects of different clustering resolutions and algorithmic implementations, eventually using FindClusters(seurat_obj, pc.use = 1:30, resolution

= 1.0, algorithm = 3, print.output = FALSE, save.SNN = TRUE) across all datasets. This method first calculates the k-nearest neighbors between cells and constructs an SNN graph, followed by cell clustering using modularity optimization. Setting FindClusters(algorithm = 3) implements the smart local moving (SLM) algorithm (Waltman and van Eck, 2013), a modularity optimization method that extends the default Louvain algorithm (Blondel et al., 2008) used within FindClusters() through the local moving heuristic (Figure S4). After obtaining all network community structures using the local moving heuristic, SLM iterates over each community and constructs a subnetwork for each. Next, the algorithm repeats the local moving heuristic to identify each of the subnetwork communities, eventually constructing a reduced network, and then repeating this process until a network can no longer be reduced (see Waltman and van Eck, 2013 for further details).

Finally, I took these clusters to generate feature plots using RunUMAP(seurat_obj, dims = 1:30, reduction = "pca"). UMAP refers to a manifold-based learning algorithm that reduces the dimensionality of highly complex data, especially in the context of biological sequencing. To construct a low-dimensional representation of the data, UMAP involves three steps. First, the algorithm constructs a weighted k-neighbor graph, followed by an edge-transformation to ambient local distance, and finally, the unification of asymmetrical topological features in the weighted k-neighbor graph (McInnes et al. 2018). The second phase of UMAP is a force-directed graph layout algorithm, which identifies a low-dimensional representation of the data to optimize an objective function (McInnes et al. 2018). Using UMAP, it is possible to visualize scRNA-seq and scATAC-seq data in two dimensions, as well as identify clusters of cells with similar gene expression and chromatin accessibility properties (Becht et al., 2018).

The R package scCustomize (Marsh et al., 2024) was used with AddModuleScore to average the expression of 332 synaptic genes for each cell, and the score of each cell was averaged per time point within each dataset. I then plotted the average synaptic gene expression level across timepoints, and observed a temporal expression profile that matches our lab's previous bulk *Drosophila* RNA-seq analysis. Genes with the highest average expression in each scRNA cluster, or gene activity scores in scATAC, were used with enriched marker genes from FindAllMarkers(seurat_obj, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25, test.use = "wilcox") to broadly annotated the RNA and ATAC clusters, which were then merged by annotation. All data processing and analysis was conducted in R (4.2.2), and scripts are available upon request. The overall processing and downstream analysis pipeline is visualized in Figure S5.

Cell Fate Trajectory Inference Using scMultiNODE

Single-cell Multi-modal Neural Ordinary Differential Equations (scMultiNODE) integrates scRNA-seq and scATAC-seq time-series data profiles through OT and neural ODEs to model cellular dynamics (Figure S6, Zhang et al., 2024). The algorithmic pseudocode implementation of scMultiNODE is depicted in Figure S7. For input, scMultiNODE takes recorded time point indices, gene expression matrices, chromatin peak/activity matrices, and hyperparameters associated with the model. The model begins by using Auto-Encoders (AEs) to compute a low-dimensional representation of each data modality:

$$\mathbf{Z}_{\mathrm{RNA}} = \mathrm{Enc}_{\mathbf{X}}(\mathbf{X}_{\mathrm{ALL}},\ \phi_{\mathbf{X}}), \quad \widehat{\mathbf{X}}_{\mathrm{ALL}} = \mathrm{Dec}_{\mathbf{X}}(\mathbf{Z}_{\mathrm{RNA}},\ \theta_{\mathbf{X}});$$
$$\mathbf{Z}_{\mathrm{ATAC}} = \mathrm{Enc}_{\mathbf{Y}}(\mathbf{Y}_{\mathrm{ALL}},\ \phi_{\mathbf{Y}}), \quad \widehat{\mathbf{Y}}_{\mathrm{ALL}} = \mathrm{Dec}_{\mathbf{Y}}(\mathbf{Z}_{\mathrm{ATAC}},\ \theta_{\mathbf{Y}}).$$

Next, the model aligns the modality-specific latent representations of the data and predicts cell correspondence between each modality using Gromov-Wasserstein (GW) OT (Peyré and Cuturi, 2019):

$$\text{GW distance} = \sum_{i,i',j,j'} \| \mathbf{D}^{\text{RNA}}_{ii'} - \mathbf{D}^{\text{ATAC}}_{jj'} \|^2 \, \mathbf{T}_{ij} \mathbf{T}_{i'j'}$$

.

In doing so, scMultiNODE is able to align cells together that exhibit similar biological activity across the two data types (Zhang et al., 2024). After alignment, scMultiNODE computes a joint latent space and intra-modality distance matrices using the fusion layer

$$\tilde{\mathbf{z}} = \text{Fus}(\mathbf{z},\ \omega) \quad \text{with} \quad \begin{cases} \mathbf{z} = \text{Enc}_{\text{RNA}}(\mathbf{x},\ \phi_{\mathbf{X}}) \quad \text{and} \quad \widehat{\mathbf{x}} = \text{Dec}_{\text{RNA}}(\tilde{\mathbf{z}},\ \theta_{\mathbf{X}}) & \text{for RNA data} \\ \mathbf{z} = \text{Enc}_{\text{ATAC}}(\mathbf{y},\ \phi_{\mathbf{Y}}) \quad \text{and} \quad \widehat{\mathbf{y}} = \text{Dec}_{\text{ATAC}}(\tilde{\mathbf{z}},\ \theta_{\mathbf{Y}}) & \text{for ATAC data} \end{cases}$$

to output low-dimensional representations of Z_{RNA} and Z_{ATAC} and then optimizes several parameters to minimize fusion loss:

$$\mathcal{L}_{\text{fusion}} = \alpha \sum_{i,j} \mathbf{1}\left(\mathbf{T}_{ij} \neq 0\right) \| \tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j \|_2^2 + \text{MSE}\left(\mathbf{X}_{\text{ALL}},\ \widehat{\mathbf{X}}_{\text{ALL}}\right) + \text{MSE}\left(\mathbf{Y}_{\text{ALL}},\ \widehat{\mathbf{Y}}_{\text{ALL}}\right) \quad \text{where}$$

$$\tilde{\mathbf{z}}_i = \text{Fus}(\mathbf{z}_i, \omega) \text{ and } \tilde{\mathbf{z}}_j = \text{Fus}(\mathbf{z}_j, \omega) \text{ for RNA cell } i \text{ and ATAC cell } j \quad \text{(from Eq. 3)}$$

$$\widehat{\mathbf{X}}_{\text{ALL}} = \text{Dec}_{\mathbf{X}}(\tilde{\mathbf{Z}}_{\text{RNA}},\ \theta_{\mathbf{X}}) \text{ with } \tilde{\mathbf{Z}}_{\text{RNA}} = \{\tilde{\mathbf{z}}_i \mid \text{all RNA cells } i\},$$

$$\widehat{\mathbf{Y}}_{\text{ALL}} = \text{Dec}_{\mathbf{Y}}(\tilde{\mathbf{Z}}_{\text{ATAC}},\ \theta_{\mathbf{Y}}) \text{ with } \tilde{\mathbf{Z}}_{\text{ATAC}} = \{\tilde{\mathbf{z}}_j \mid \text{all ATAC cells } j\}.$$

Finally, the entire model is optimized to minimize a dynamic loss function

$$\mathcal{L}_{\text{dyn}} = \sum_{t \in \mathcal{T}_{\text{RNA}}} \text{Wass}\left(\mathbf{X}^{(t)},\ \widehat{\mathbf{X}}^{(t)}\right) + \sum_{t \in \mathcal{T}_{\text{ATAC}}} \text{Wass}\left(\mathbf{Y}^{(t)},\ \widehat{\mathbf{Y}}^{(t)}\right) + \beta \mathcal{R}\left(\mathcal{T}_{\text{RNA}}, \mathcal{T}_{\text{ATAC}}\right),$$

thereby ensuring the capture of relevant cellular dynamics.

To construct cell fate transition pathways, scMultiNODE uses the LAP method, finding an optimal path between cell states while simultaneously minimizing action and transition time. Given starting point $\mathbf{X}_0$ (cell representations in joint latent space at t=0) and end point $\mathbf{X}_K$, LAP finds a path represented as a sequence of K points $\mathscr{P} = \{\mathbf{X}_0, \ldots, \mathbf{X}_K\}$. The tangential velocity for

each segment between $\mathbf{X}_{k-1}$ and $\mathbf{X}_k$ is defined as $V_k = \frac{(X_k - X_{k-1}}{\Delta}$, where Δ represents the time step taken by cells from $\mathbf{X}_{k-1}$. The action $\mathscr{S}$ on path $\mathscr{P}$ is, therefore:

$$\mathcal{S} = \frac{1}{2}\sum_{k=1}^{K}\left(\mathbf{V}_k - \text{Drift}_{\omega}(\tilde{\mathbf{X}}_k)\right)^2 \Delta \quad \text{with } \tilde{\mathbf{X}}_k = \frac{\mathbf{X}_{k-1} + \mathbf{X} - k}{2}.$$

LAP seeks to align V_k with the differential velocity $\text{Drift}_{\omega}(\tilde{X}_k)$ learned by scMultiNODE, minimizing transition time. The optimal path is then the objective function

$$\widehat{\mathcal{P}} = \underset{\mathcal{P},\Delta}{\text{argmin}}\,\mathcal{S} = \underset{\mathcal{P},\Delta}{\text{argmin}}\,\frac{1}{2}\sum_{k=1}^{K}\left(\mathbf{V}_k - \text{Drift}(\tilde{\mathbf{X}}_k, \omega)\right)^2 \Delta$$

and solved through two interable steps. For a more comprehensive description of scMultiNODE, please see Zhang et al., 2024.

3.4.2 Data availability

scRNA-seq data

For Uzquiano et al. (2022), the scRNA-seq read-level data is available at https://www.synapse.org/Synapse:syn26346373/files/. The count-level data and metadata are available at the Single Cell Portal under the study code SCP1756. For Sebastian et al. (2023), the scRNA-seq read-level data is available at NCBI GEO under accession code #GSE228315. For Zenk et al. (2024), the raw sequencing data is available from the European Genome Phenome Archive (https://ega-archive.org/studies/EGAS50000000155). The processed data is accessible at https://doi.org/10.5281/zenodo.10471808.

Paired multi omics data (scRNA-seq & scATAC-seq)

For Fleck et al. (2022) the raw sequencing data is available at ArrayExpress: E-MTAB-12001 for developmental time course scRNA-seq data, E-MTAB-11998 for

developmental time course scATAC-seq data, and E-MTAB-12004 for neuroepithelial stage multi-omics data. For Zhu et al. (2023), the raw and processed snRNA-seq and snATAC-seq data are deposited at NCBI GEO under accession code #GSE204684.

3.4.3 Code availability & resources

The human cortex dataset and code for scMultiNODE are publicly available at https://github.com/rsinghlab/scMultiNODE. Custom scripts in R (4.2.2) for all other single-cell data preprocessing and analysis are available upon request. All scripts were executed using an integrated RStudio Server with 64 cores and 492 GB of memory. This research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University.

3.5 Figures

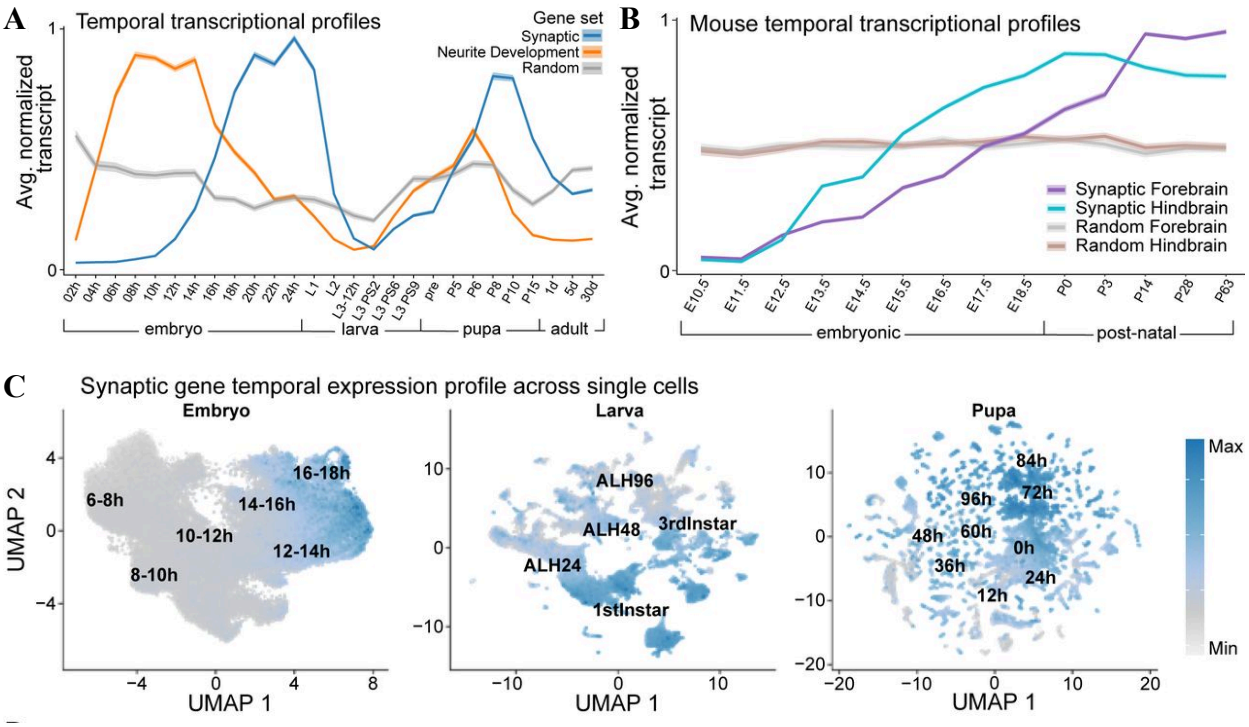


Figure 12: Broad temporal coordination of synaptic gene expression in *Drosophila* and mouse. (A) Average temporal expression profile of the 485 genes most highly correlated with synaptic expression profile footprint (blue) is distinct from the average expression profile of 485 genes that regulate neurite development (orange) or a random set of nervous system expressed genes (grey) (Graveley et al., 2011). **(B)** Average profile of 742 mouse genes temporally correlated with the expression pattern of orthologs of the 20 *Drosophila* synaptic footprint genes shows coordinated upregulation around embryonic day 12 (Cardoso-Moreira et al., 2019). **(C)** Average expression scores of the 485 coordinately expressed synaptic genes across time-series scRNA-seq datasets (Calderon et al., 2022; Corrales et al., 2022; Kurmangaliyev et al., 2020) displayed as UMAP feature plots yield a similar temporal expression pattern. Adapted from Kentro et al. (2024).

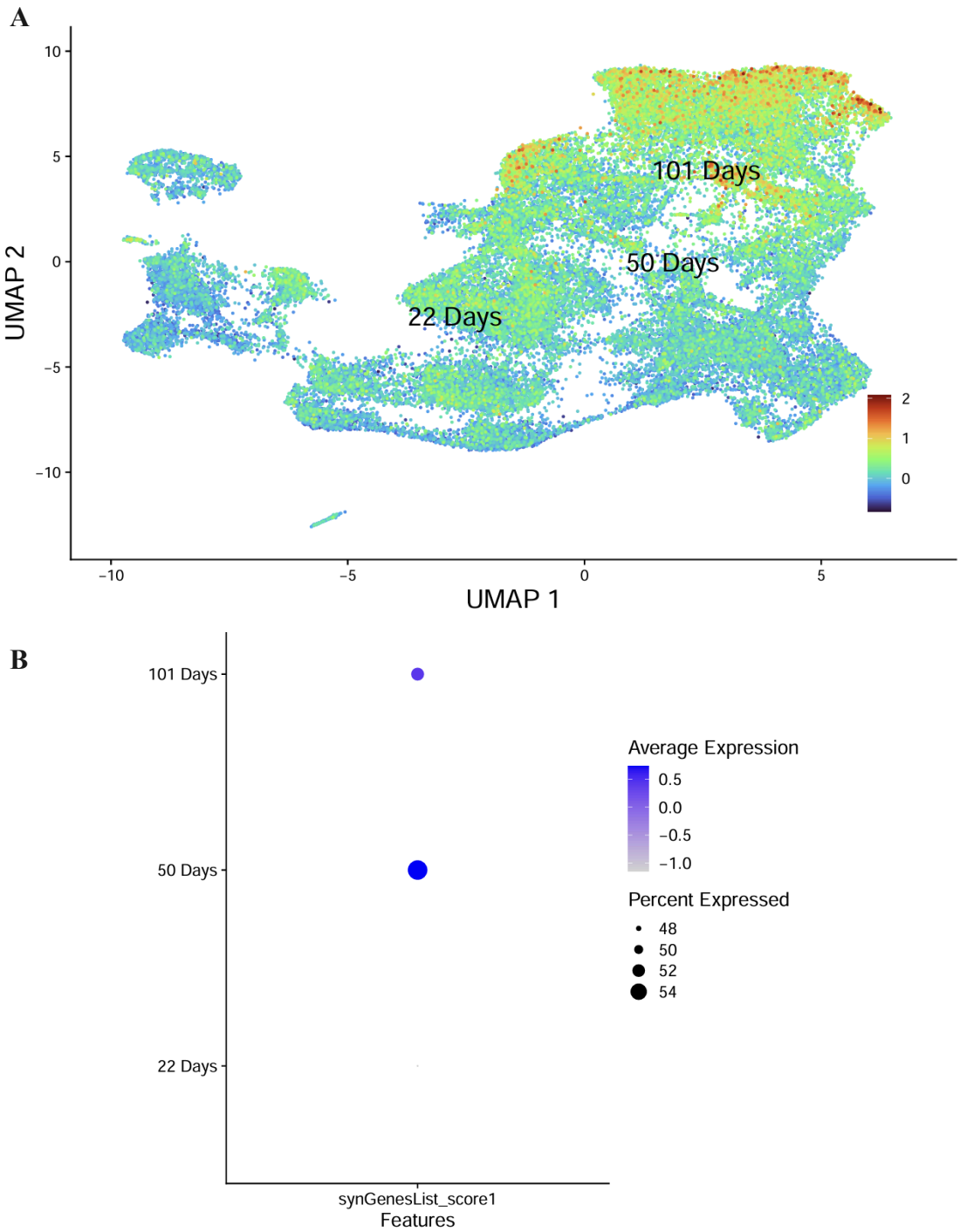


Figure 13: Average expression scores of the 332 coordinately expressed synaptic gene orthologs across a day 22 to 101 time-series scRNA-seq dataset. (A) Feature plot of average synaptic gene expression scores. **(B)** Dotplot of synaptic gene expression averaged per time point The visualizations yielded similar temporal expression patterns across development. Dataset from Sebastian et al. (2023).

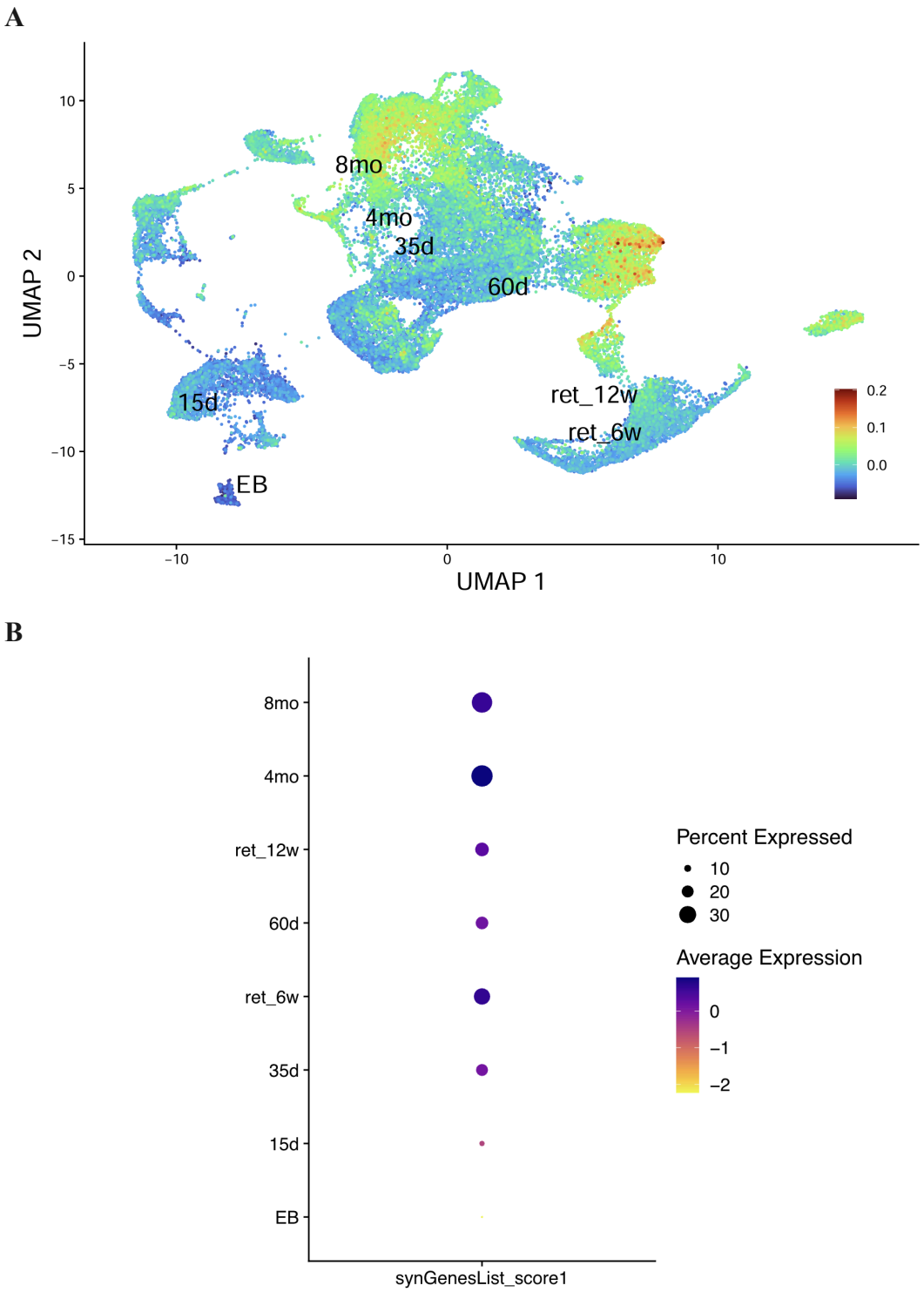


Figure 14: Average expression scores of the 332 coordinately expressed synaptic gene orthologs across an EB to 8 months time-series scRNA-seq dataset. (A) Feature plot of average synaptic gene expression scores. **(B)** Dotplot of synaptic gene expression averaged per time point, omitting EB and 63d. The visualizations yielded similar synaptic temporal expression patterns across development. Dataset from Zenk et al. (2024).

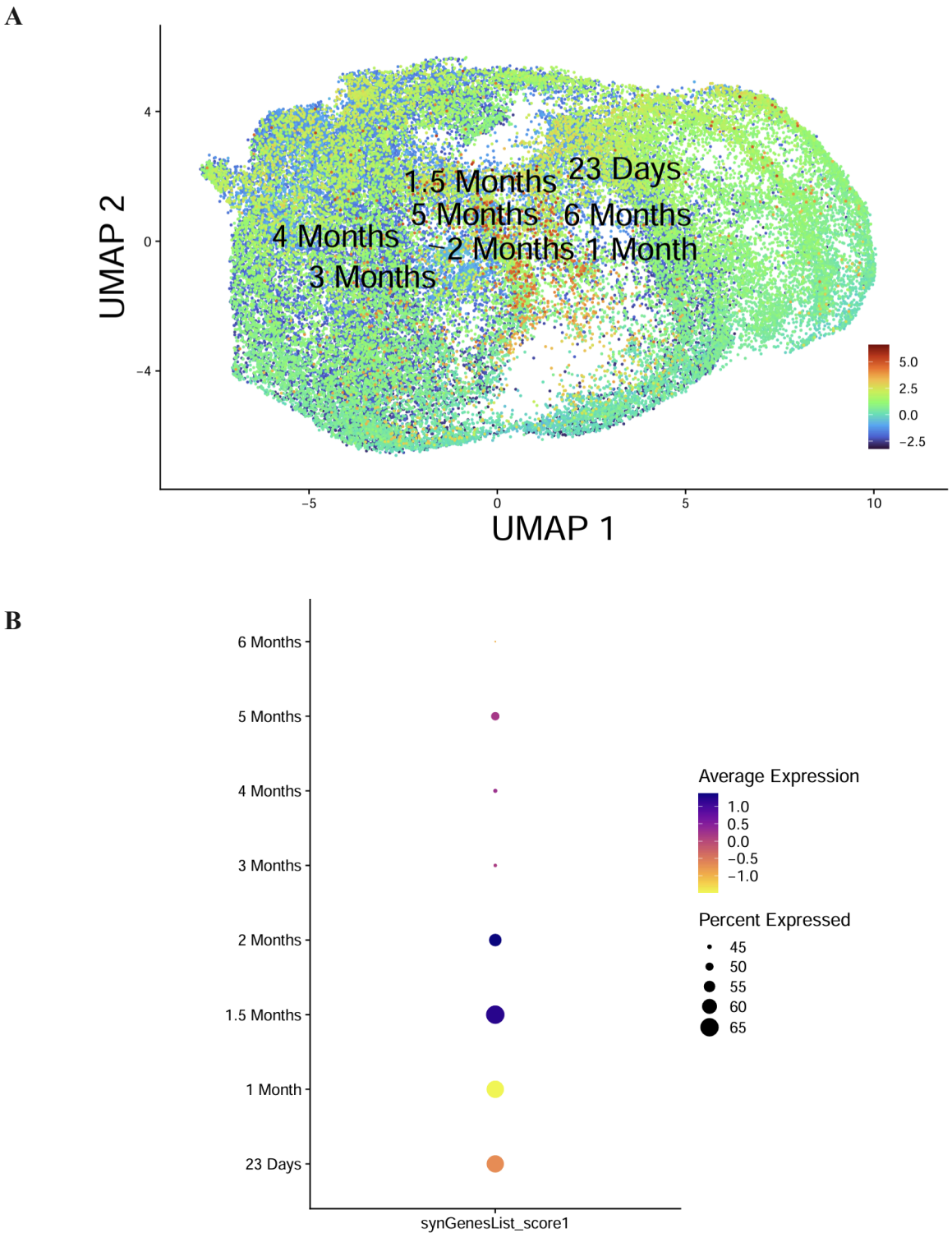
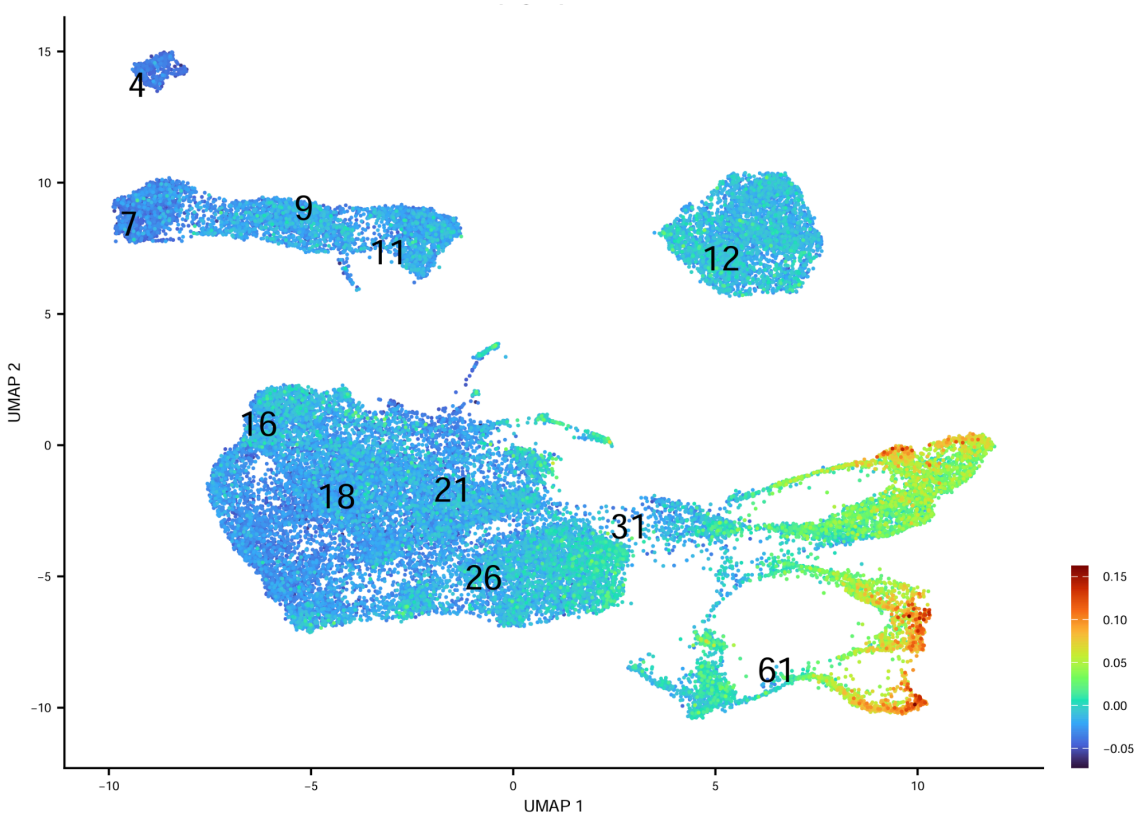
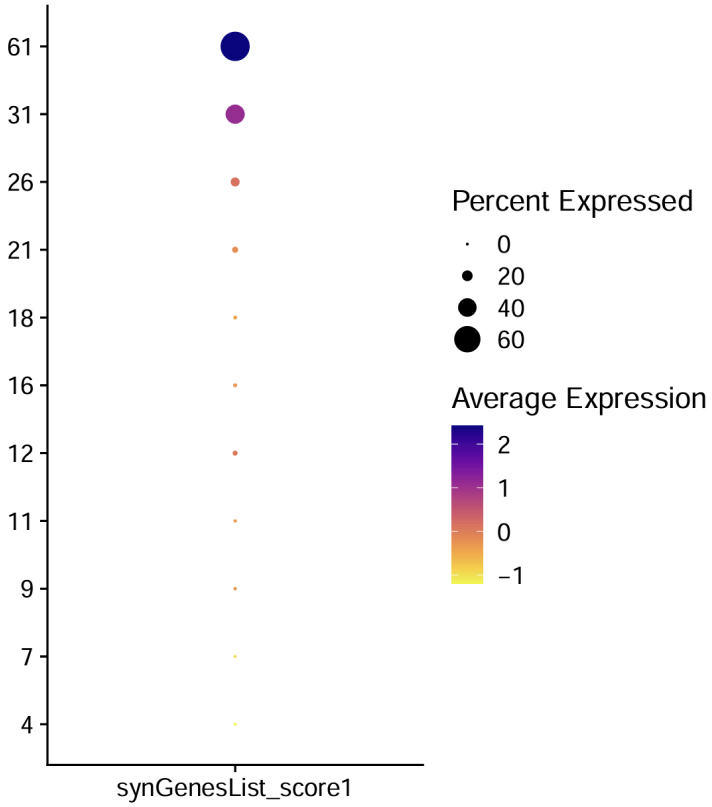


Figure 15: Average expression scores of the 332 coordinately expressed synaptic gene orthologs across a 23 days to 6 months time-series scRNA-seq dataset. (A) Feature plot of average synaptic gene expression scores. **(B)** Dotplot of synaptic gene expression, averaged per time point. Dataset from Uzquiano et al. (2022).

A


4
7
9
11
12
16
18
21
26
31
61
UMAP 1
UMAP 2
15
10
5
0
−5
−10
−10
−5
0
5
10
0.15
0.10
0.05
0.00
−0.05

B


61
31
26
21
18
16
12
11
9
7
4
synGenesList_score1
Percent Expressed
0
20
40
60
Average Expression
2
1
0
−1

C

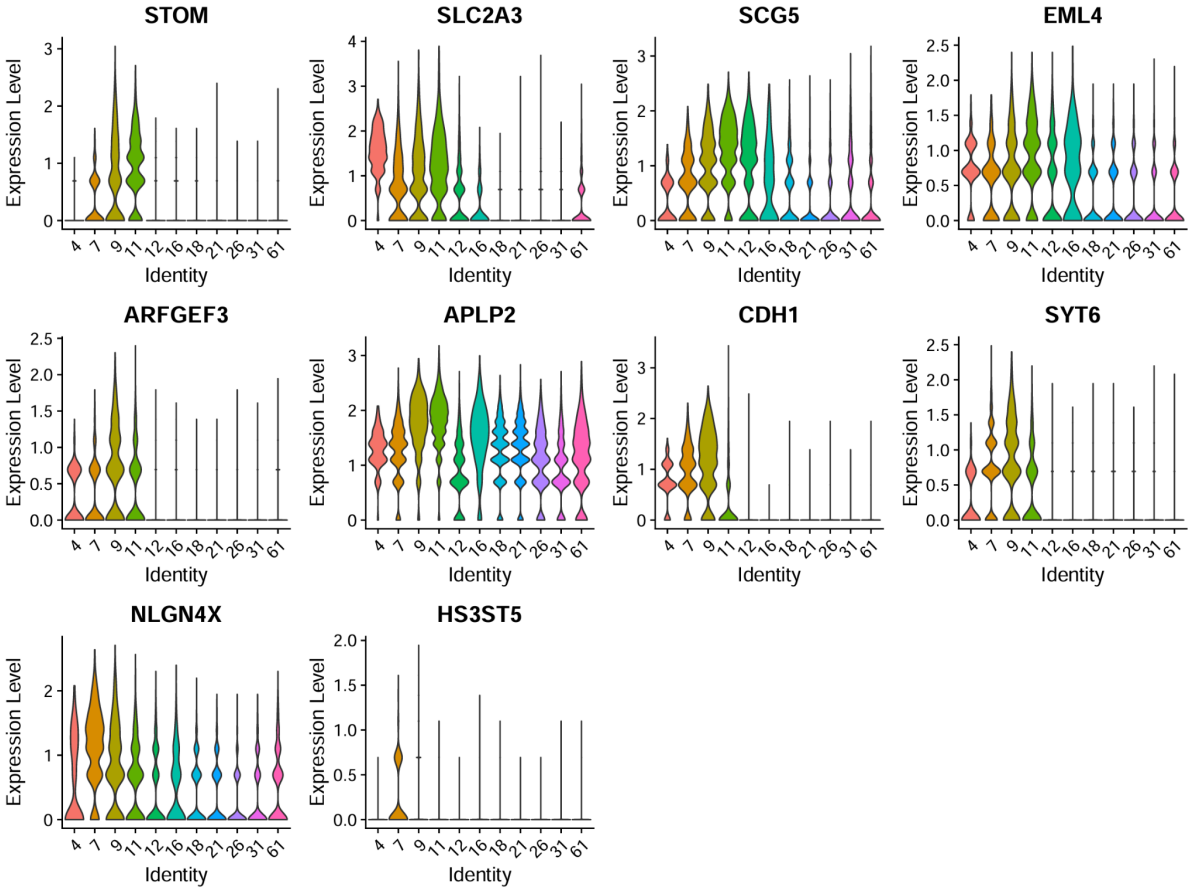
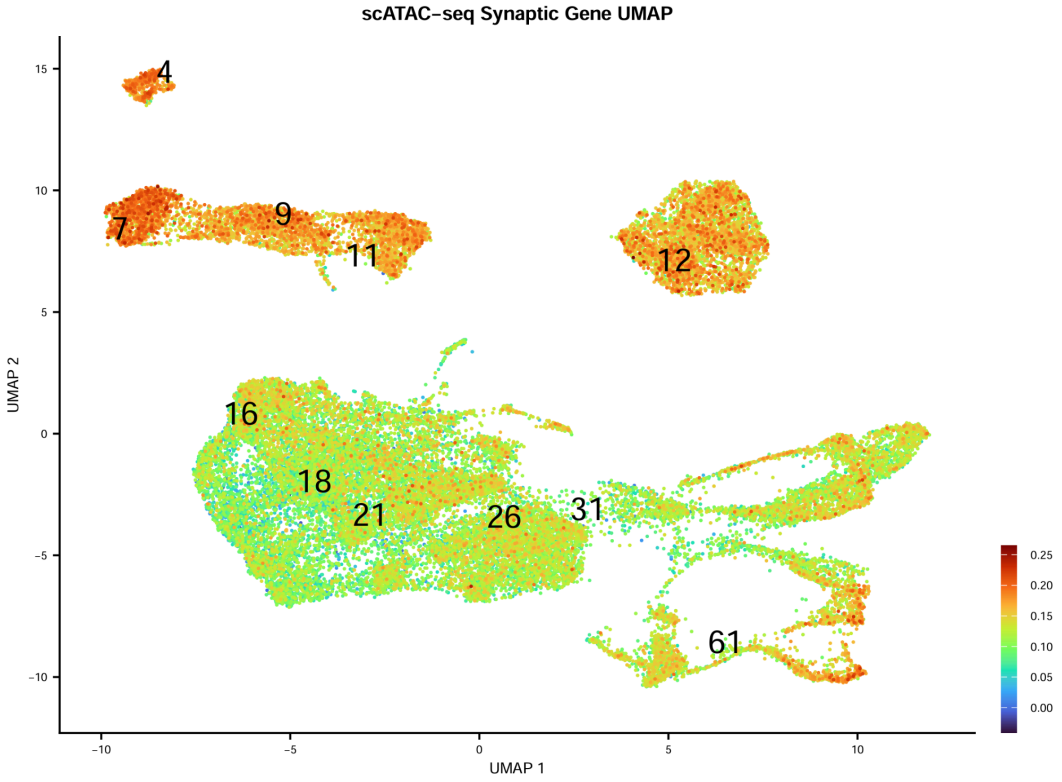
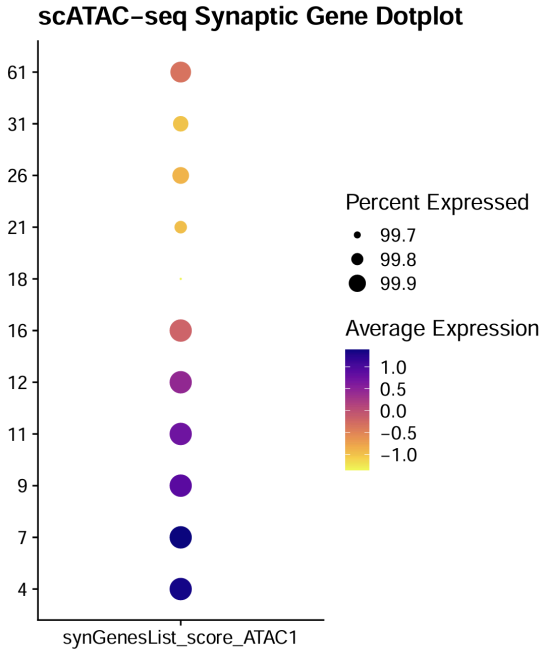


Figure 16: Average expression scores of the 332 coordinately expressed synaptic gene orthologs across a day 4 to 61 time-series scRNA-seq dataset. (A) Feature plot of average synaptic gene expression scores and **(B)** dotplot of data, averaged per time point. The visualizations yielded similar temporal expression patterns across development. **(C)** Violin plots of the top 10 globally expressed synaptic genes. Dataset from Fleck et al. (2022).

A


scATAC-seq Synaptic Gene UMAP
UMAP 2
UMAP 1
4
7
9
11
12
16
18
21
26
31
61
15
10
5
0
−5
−10
−10
−5
0
5
10
0.25
0.20
0.15
0.10
0.05
0.00

B


scATAC-seq Synaptic Gene Dotplot
61
31
26
21
18
16
12
11
9
7
4
synGenesList_score_ATAC1
Percent Expressed
99.7
99.8
99.9
Average Expression
1.0
0.5
0.0
−0.5
−1.0

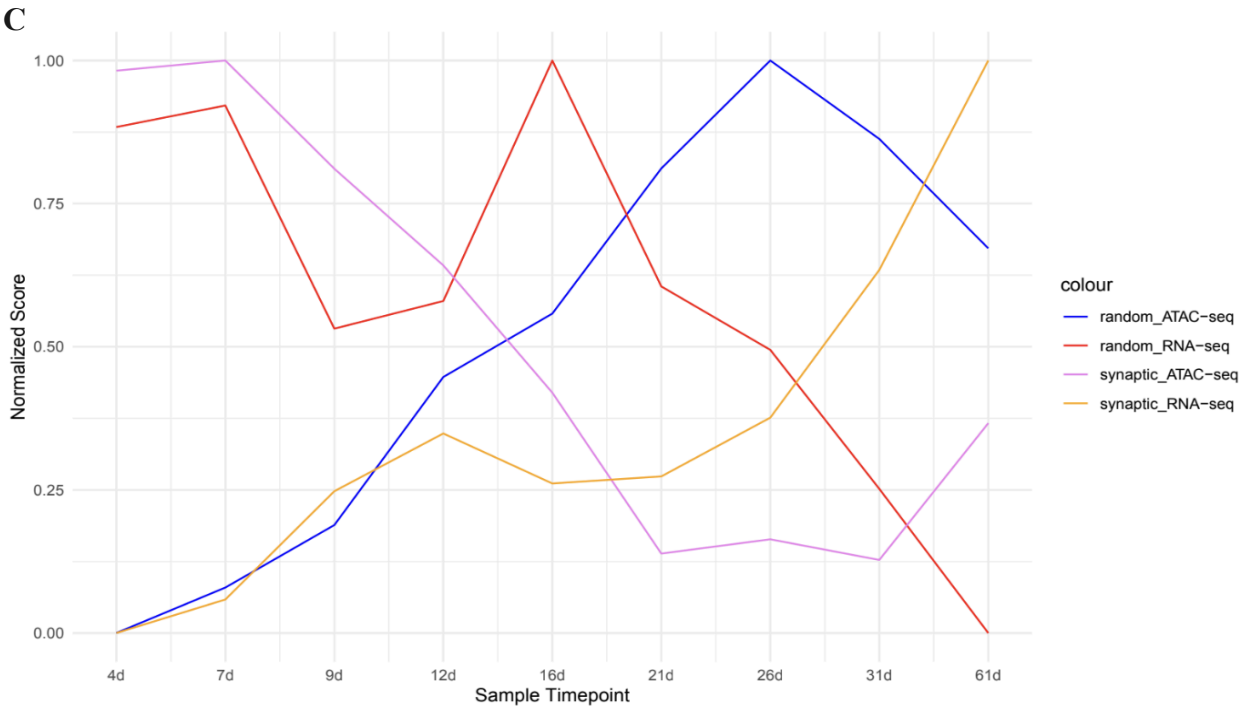


Figure 17: Average chromatin accessibility scores of the 332 coordinately expressed synaptic gene orthologs across a day 4 to 61 time-series scATAC-seq dataset. (A) Feature plot of average synaptic chromatin accessibility scores and **(B)** dotplot of data, averaged per time point. The visualizations yielded similar temporal accessibility patterns across development. **(C)** Normalized module score at each developmental timepoint, comparing random gene chromatin accessibility (blue line), random gene expression (red line), synaptic gene chromatin accessibility (pink line), and synaptic gene expression (yellow line). Time points 11d and 18d omitted for visual clarity.

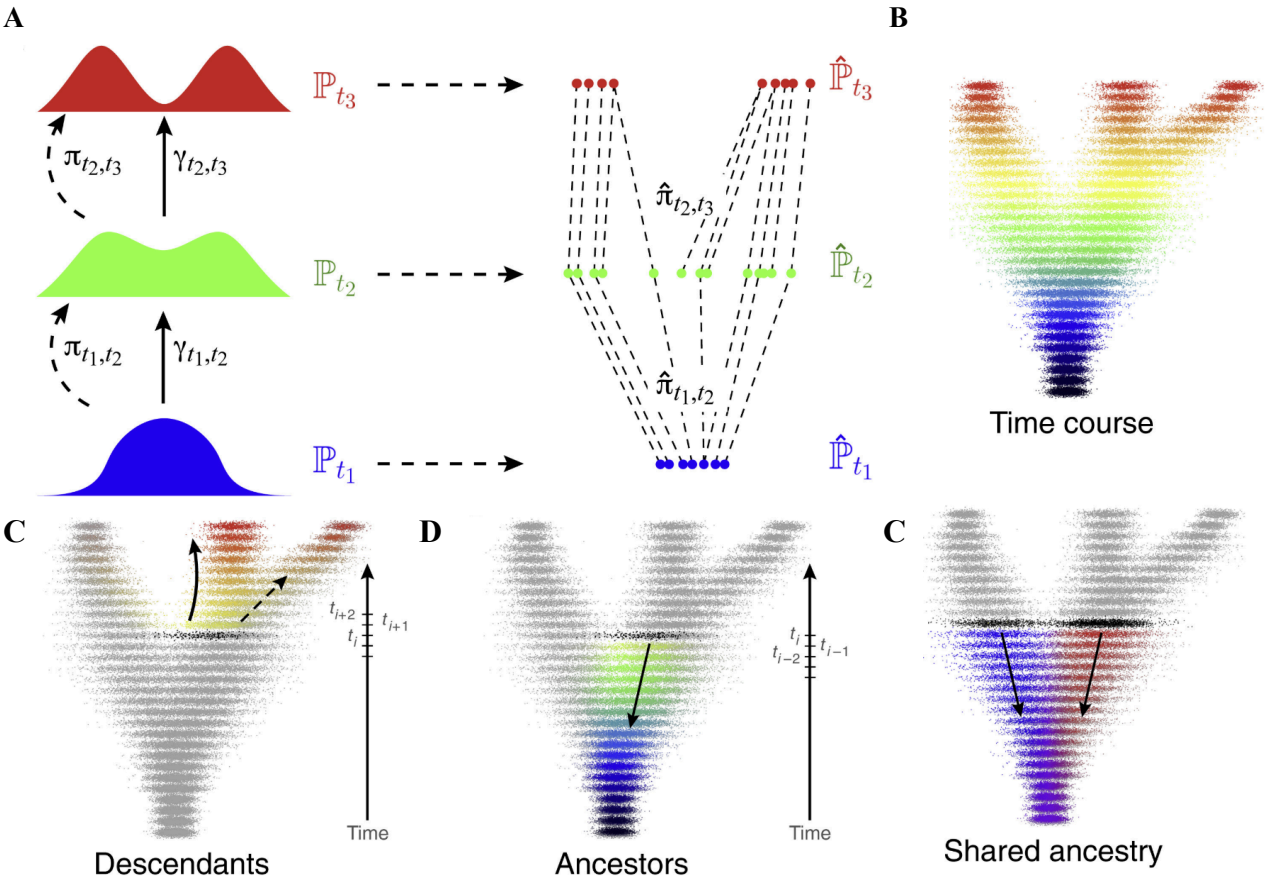


Figure 18: Example OT modeling of cell developmental process at various time points t_1, t_2, and t_3. (A) Temporal progression of a time-varying distribution $\mathbb{P}_t$ (left) can be sampled to obtain distributions of cells $\widehat{\mathbb{P}}_{t_i}$. (right) at each time point. Here, $\gamma_{t1\ ,\ t2}$ represents the unknown true coupling per cell, which is assumed to be close to the optimal transport coupling ($\pi_{t1\ ,\ t2}$), approximated by $\widehat{\pi}_{t1\ ,\ t2}$ from the empirical distributions $\widehat{\mathbb{P}}_{t1}$ and $\widehat{\mathbb{P}}_{t2}$. **(B)** Single-cell gene expression profiles (individual dots) over developmental time course, colored by collection time. **(C)** Descendants of a cell set at later time points. **(D)** Ancestors of a cell set at earlier time points. **(E)** Shared ancestry of two cell sets (black dots), where ancestors of each cell population are blue and red, respectively. Each dot represents a single-cell. Adapted from Schiebinger et al. (2019).

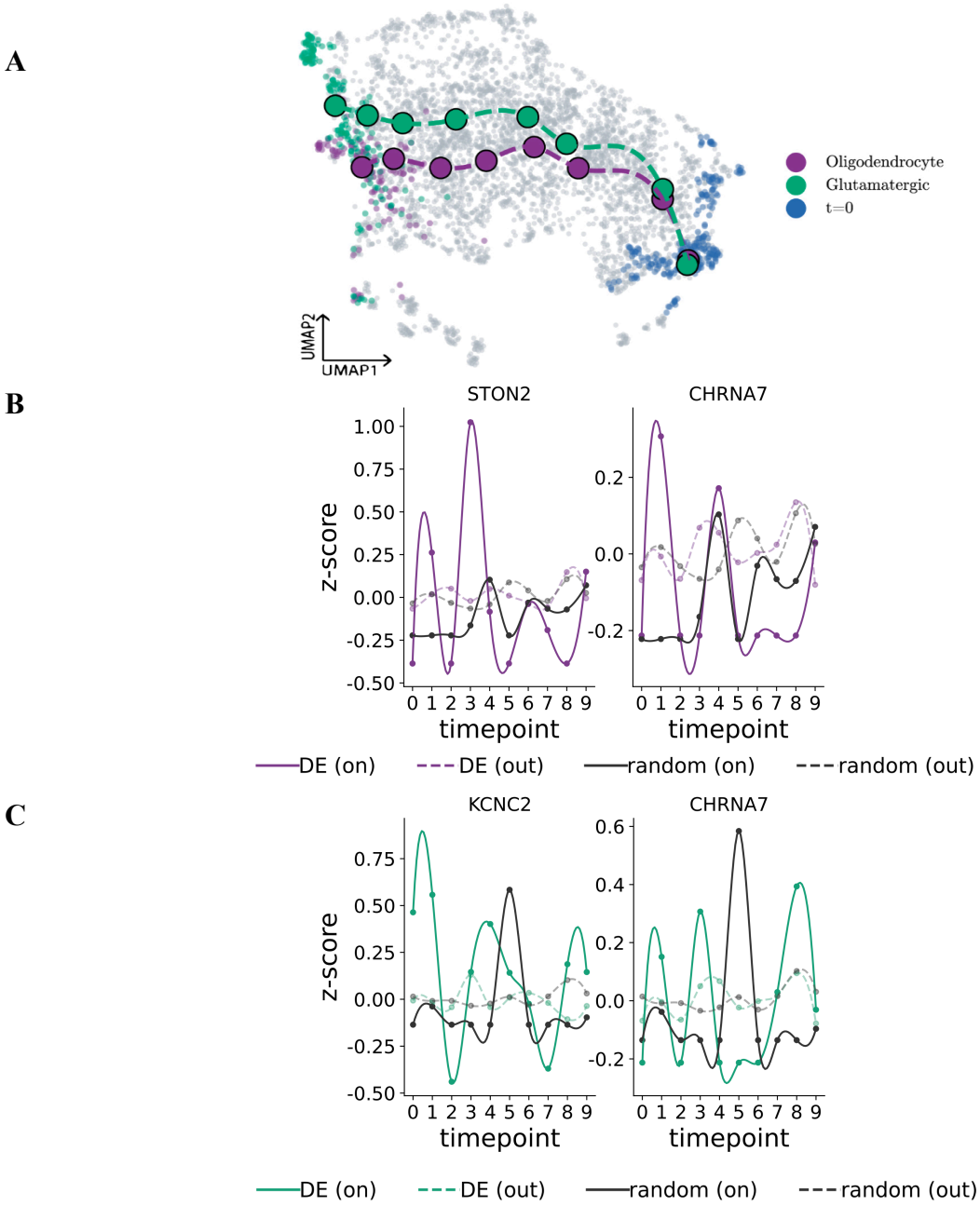


Figure 19: scMultiNODE trajectory inference of oligodendrocytes and glutamatergic neurons in human cerebral cortex. (A) 2D UMAP plot of least action path (LAP) from cells at t = 0 to oligodendrocyte and glutamatergic neuron populations (Zhang et al., 2024). **(B)** Gene expression z-score values of differentially expressed genes STON2 and CHRNA7 on the oligodendrocyte path. **(C)** Gene expression z-score values of differentially expressed genes KCNC2 and CHRNA7 on the glutamatergic neuron path. Colored lines represent differentially expressed gene z-scores per timepoint, averaged for cells on the path (solid) and off the path (dotted). Black lines represent the average z-score of five random genes for cells on the path (solid) and off the path (dotted). Human cerebral cortex dataset taken from Zhu et al. (2023).

SmartLocalMovingAlgorithm

input:

A: Adjacency matrix of a network

c: Initial assignment of nodes to communities

output:

c: Final assignment of nodes to communities

// Run the local moving heuristic.

$c \leftarrow$ LocalMovingHeuristic(A, c)

if NumberOfCommunities(c) < NumberOfNodes(A) **then**

// For each community, construct a subnetwork and run the local moving heuristic.

// Construct a reduced network based on the community structure of the subnetworks.

$c_{old} \leftarrow c$

$j \leftarrow 0$

for $i \leftarrow 1$ **to** NumberOfCommunities(c_{old}) **do**

$A_{sub} \leftarrow$ Subnetwork(A, c_{old}, i)

$c_{sub} \leftarrow$ [1…NumberOfNodes(A_{sub})]

$c_{sub} \leftarrow$ LocalMovingHeuristic(A_{sub}, c_{sub})

$c(c_{old} = i) \leftarrow c_{sub} + j$

$c_{reduced}([j + 1]…[j +$ NumberOfCommunities(c_{sub})]) $\leftarrow i$

$j \leftarrow j +$ NumberOfCommunities(c_{sub})

end for

$A_{reduced} \leftarrow$ ReducedNetwork(A, c)

// Perform a recursive call to identify the community structure of the reduced network.

$c_{reduced} \leftarrow$ SmartLocalMovingAlgorithm($A_{reduced}$, $c_{reduced}$)

// Merge communities based on the community structure of the reduced network.

$c_{old} \leftarrow c$

for $i \leftarrow 1$ **to** NumberOfCommunities(c_{old}) **do**

$c(c_{old} = i) \leftarrow c_{reduced}(i)$

end for

end if

Figure S4: SLM algorithm pseudocode. Adapted from Waltman and van Eck (2013).

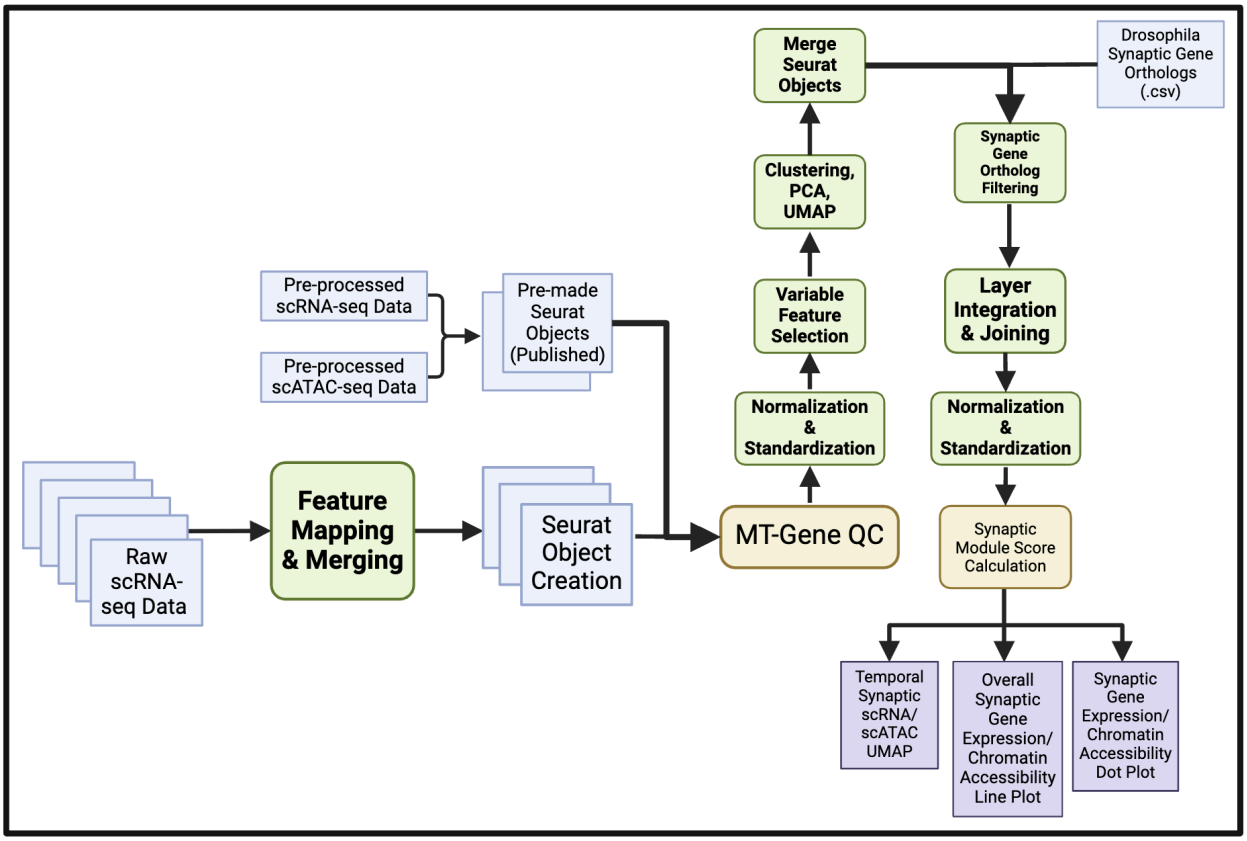


Figure S5: Single-cell human data processing and downstream analysis pipeline. Raw scRNA-seq reads were merged across several developmental time points and then converted to a Seurat object. Both created and pre-made Seurat objects underwent QC metrics, normalization and standardization, variable feature selection, and clustering. After filtering for high and mid-matching synaptic gene orthologs, Seurat object layers were integrated and joined together, re-normalized and standardized, and then used to calculate a synaptic module score for downstream visualization and analysis. Created with Biorender.com.

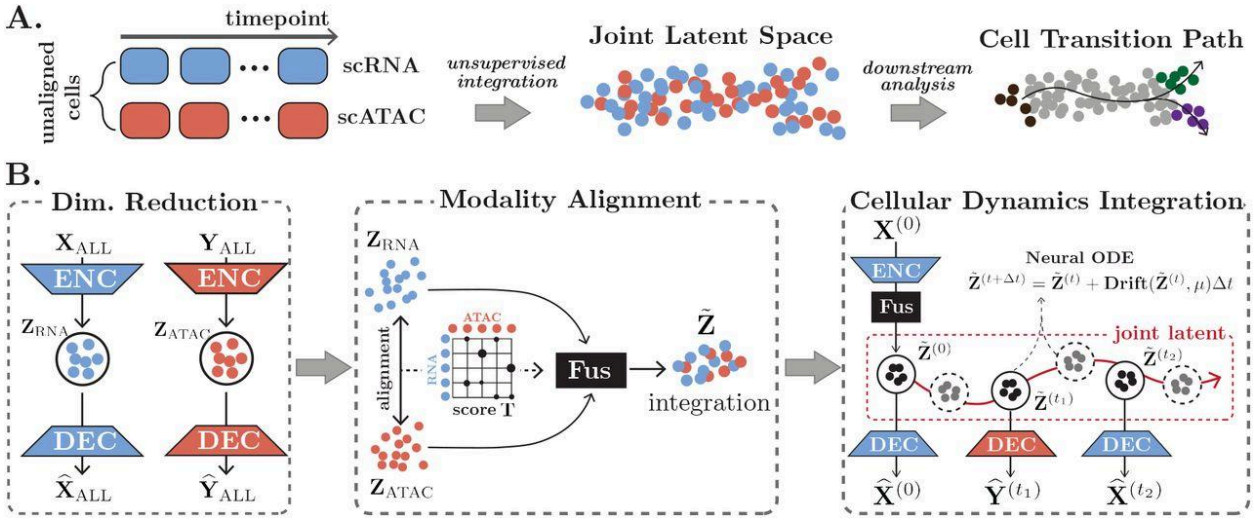


Figure S6: Representative scMultiNODE model architecture and overview. (A) scMultiNODE assumes no cell-to-cell correspondence across modalities or timepoints, taking gene expression and chromatin peak matrices as input. Unsupervised learning allows scMultiNODE to create a joint latent space between data modalities, and then allow for downstream cell transition or developmental trajectory inference analysis. **(B)** scMultiNODE uses autoencoders (AEs) to identify an optimal latent space representation of each data modality, and then aligns latent representations with GW optimal transport, thereby predicting cell correspondence between the modalities. Next, the model constructs a joint latent space by mapping each modality-specific latent representation, using the predicted correspondence in which aligned cells stay together. Finally, scMultiNODE solves neural ordinary differential equations (ODEs) to adjust the optimal joint latent space to match the population level cellular dynamic by embedding dynamics in the joint latent space. Adapted from Zhang et al., 2024.

Input: The set of timepoint indices $\mathcal{T}_{\text{RNA}}$ and $\mathcal{T}_{\text{ATAC}}$; gene expression matrices $\{\mathbf{X}^{(t)} \mid t \in \mathcal{T}_{\text{RNA}}\}$; chromatin peak/gene activity matrices $\{\mathbf{Y}^{(t)} \mid t \in \mathcal{T}_{\text{ATAC}}\}$; hyperparameters Δt, α, β, the number of neighbors k in kNN; randomly initialized neural networks $\text{Enc}_{\mathbf{X}}$, $\text{Enc}_{\mathbf{Y}}$, $\text{Dec}_{\mathbf{X}}$, $\text{Dec}_{\mathbf{Y}}$, fusion layer Fus, and the neural ODE drift network Drift.

(Step I: dimensionality reduction)

$\mathbf{X}_{\text{ALL}} = \text{CONCAT}\left(\mathbf{X}^{(t)} \mid t \in \mathcal{T}_{\text{RNA}}\right)$ // concatenate cells from RNA-seq measurements

$\mathbf{Y}_{\text{ALL}} = \text{CONCAT}\left(\mathbf{Y}^{(t)} \mid t \in \mathcal{T}_{\text{ATAC}}\right)$ // concatenate cells from ATAC-seq measurements

Optimize RNA-related AE parameters $\phi_{\mathbf{X}}$ and $\theta_{\mathbf{X}}$ to minimize $\mathcal{L}_{\text{RNA}}$ (Eq. 2)

Optimize ATAC-related AE parameters $\phi_{\mathbf{Y}}$ and $\theta_{\mathbf{Y}}$ to minimize $\mathcal{L}_{\text{ATAC}}$ (Eq. 2)

(Step II: modality integration)

Construct intra-modality distance matrices $\mathbf{D}^{\text{RNA}}$ and $\mathbf{D}^{\text{ATAC}}$ with representations $\mathbf{Z}_{\text{RNA}}$ and $\mathbf{Z}_{\text{ATAC}}$

Use Quantized Gromov-Wasserstein (QGW) algorithm [11] to predict cell correspondence matrix $\mathbf{T}$ (Eq. 4)

Map latent representations to joint latent space through fusion layer (Eq. 3)

Optimize fusion layer parameter (ω) and AE decoder parameters ($\theta_{\mathbf{X}}$ and $\theta_{\mathbf{Y}}$) to minimize $\mathcal{L}_{\text{fusion}}$ (Eq. 5)

(Step III: cellular dynamics)

Optimize the entire model to minimize $\mathcal{L}_{\text{dyn}}$ (Eq. 10)

Output: modality integration $\tilde{\mathbf{Z}}_{\text{RNA}} = \text{Fus}(\mathbf{Z}_{\text{RNA}}, \omega)$ and $\tilde{\mathbf{Z}}_{\text{ATAC}} = \text{Fus}(\mathbf{Z}_{\text{ATAC}}, \omega)$

Figure S7: scMultiNODE algorithm pseudo-code. After an input set of timepoint indices, gene expression matrices, and chromatin peal/gene activity matrices, scMultiNODE performs three sequential steps: dimensionality reduction, modality integration, and the modeling of cellular dynamics. Adapted from Zhang et al., 2024.

3.6 Tables

TPs	22d (Rep #1)	22d (Rep #2)	50d (Rep #1)	50d (Rep #2)	101d (Rep #1)	101d (Rep #2)
# cells (per replicate)	8899	9071	7584	5897	7288	9851
# cells (combined)	17970		13481		17139	

Table 3: Number of analyzed cells across developmental time points from Sebastian et al. (2023) dataset. Replicate is abbreviated as Rep, and day is abbreviated as d.

TPs	EB	15d	35d	ret_6w	60d	ret_12w	4mo	8mo
Cell #	410	3534	5830	3771	8935	3385	3632	8652

Table 4: Number of analyzed organoid cells across developmental time points from Zenk et al. (2024) dataset. EB refers to pluripotent embryoid bodies recorded at day 5, day is abbreviated as d, and month is abbreviated as mo. Ret refers to retinal organoids, while all time points are taken from cerebral organoids.

TPs	23d	1mo	1.5mo	2mo	3mo	4mo	5mo	6mo
Cell #	29736	71372	38574	39299	159533	38348	31787	123765
Reduced Cell #	2974	7138	3858	3930	15954	3835	3179	12377

Table 5: Number of analyzed cells across developmental time points from Uzquiano et al. (2022) dataset. Following data preprocessing with all cells, 10% of cells were randomly sampled for downstream analysis at each timepoint. Day is abbreviated as d, while month is abbreviated as mo.

TPs	4d	7d	9d	11d	12d	16d	18d	21d	26d	31d	61d
Cell #	501	1309	1630	1560	4100	361	5842	5721	4538	5181	3345

Table 6: Number of analyzed cells across developmental time points from Fleck et al. (2022) dataset. Day is abbreviated as d.

TPs	GW-18	GW-19	GW-23	GW-24	<1 yo	4yo	6yo	14yo	20yo	39yo
Cell #	2801	3679	3102	4754	6406	5418	3693	8300	1541	5855
Reduced Cell #	140	184	155	238	320	271	185	415	77	293

Table 7: Number of analyzed human cerebral cortex cells across developmental time points from Zhu et al. (2022) dataset. Abbreviations are gestational week (GW), and year (yo). Following data preprocessing with all cells, 5% of cells were randomly sampled for downstream analysis at each timepoint.

3.7 References

Ancín, I., J. A. Cabranes, J. L. Santos, E. Sánchez-Morla, B. Vázquez-Álvarez, L. Rodríguez-Moya, A. Pousada-Casal, C. Fernández, A. Aparicio, and A. Barabash. 2011. "*CHRNA7* Haplotypes Are Associated with Impaired Attention in Euthymic Bipolar Disorder." *Journal of Affective Disorders* 133 (1): 340–45. https://doi.org/10.1016/j.jad.2011.04.008.

Becht, Etienne, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W. H. Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W. Newell. 2019. "Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP." *Nature Biotechnology* 37 (1): 38–44. https://doi.org/10.1038/nbt.4314.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10): P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

Bruer, John T. 1999. *The Myth of the First Three Years: A New Understanding of Early Brain Development and Lifelong Learning*. Simon and Schuster.

Calderon, Diego, Ronnie Blecher-Gonen, Xingfan Huang, Stefano Secchia, James Kentro, Riza M. Daza, Beth Martin, et al. 2022. "The Continuum of Drosophila Embryonic Development at Single-Cell Resolution." *Science* 377 (6606): eabn5800. https://doi.org/10.1126/science.abn5800.

Cardoso-Moreira, Margarida, Jean Halbert, Delphine Valloton, Britta Velten, Chunyan Chen, Yi Shao, Angélica Liechti, et al. 2019. "Gene Expression across Mammalian Organ Development." *Nature* 571 (7766): 505–9. https://doi.org/10.1038/s41586-019-1338-5.

Clare, Ryan, Victoria G. King, Martin Wirenfeldt, and Harry V. Vinters. 2010. "Synapse Loss in Dementias." *Journal of Neuroscience Research* 88 (10): 2083–90. https://doi.org/10.1002/jnr.22392.

Corrales, Marc, Benjamin T. Cocanougher, Andrea B. Kohn, Jason D. Wittenbach, Xi S. Long, Andrew Lemire, Albert Cardona, Robert H. Singer, Leonid L. Moroz, and Marta Zlatic. 2022. "A Single-Cell Transcriptomic Atlas of Complete Insect Nervous Systems across Multiple Life Stages." *Neural Development* 17 (1): 8. https://doi.org/10.1186/s13064-022-00164-6.

Court, J., D. Spurden, S. Lloyd, I. McKeith, C. Ballard, N. Cairns, R. Kerwin, R. Perry, and E. Perry. 1999. "Neuronal Nicotinic Receptors in Dementia with Lewy Bodies and Schizophrenia: Alpha-Bungarotoxin and Nicotine Binding in the Thalamus." *Journal of Neurochemistry* 73 (4): 1590–97. https://doi.org/10.1046/j.1471-4159.1999.0731590.x.

Davies, C A, D M Mann, P Q Sumpter, and P O Yates. 1987. "A Quantitative Morphometric Analysis of the Neuronal and Synaptic Content of the Frontal and Temporal Cortex in Patients with Alzheimer's Disease." *Journal of the Neurological Sciences* 78 (2): 151–64. https://doi.org/10.1016/0022-510x(87)90057-8.

Fleck, Jonas Simon, Sophie Martina Johanna Jansen, Damian Wollny, Fides Zenk, Makiko Seimiya, Akanksha Jain, Ryoko Okamoto, et al. 2023. "Inferring and Perturbing Cell Fate Regulomes in Human Brain Organoids." *Nature* 621 (7978): 365–72. https://doi.org/10.1038/s41586-022-05279-8.

Freedman, Robert, Michael Hall, Lawrence E. Adler, and Sherry Leonard. 1995. "Evidence in Postmortem Brain Tissue for Decreased Numbers of Hippocampal Nicotinic Receptors in Schizophrenia." *Biological Psychiatry* 38 (1): 22–33. https://doi.org/10.1016/0006-3223(94)00252-X.

Goyal, Manu S., and Marcus E. Raichle. 2013. “Gene Expression-Based Modeling of Human Cortical Synaptic Density.” *Proceedings of the National Academy of Sciences* 110 (16): 6571–76. https://doi.org/10.1073/pnas.1303453110.

Graveley, Brenton R., Angela N. Brooks, Joseph W. Carlson, Michael O. Duff, Jane M. Landolin, Li Yang, Carlo G. Artieri, et al. 2011. “The Developmental Transcriptome of Drosophila Melanogaster.” *Nature* 471 (7339): 473–79. https://doi.org/10.1038/nature09715.

Hafemeister, Christoph, and Rahul Satija. 2019. “Normalization and Variance Stabilization of Single-Cell RNA-Seq Data Using Regularized Negative Binomial Regression.” *Genome Biology* 20 (1): 296. https://doi.org/10.1186/s13059-019-1874-1.

Hao, Yuhan, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck, Shiwei Zheng, Andrew Butler, Maddie J. Lee, et al. 2021. “Integrated Analysis of Multimodal Single-Cell Data.” *Cell* 184 (13): 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

Hu, Yanhui, Aram Comjean, Jonathan Rodiger, Yifang Liu, Yue Gao, Verena Chung, Jonathan Zirin, Norbert Perrimon, and Stephanie E Mohr. 2021. “FlyRNAi.Org—the Database of the Drosophila RNAi Screening Center and Transgenic RNAi Project: 2021 Update.” *Nucleic Acids Research* 49 (D1): D908–15. https://doi.org/10.1093/nar/gkaa936.

Huttenlocher, Peter R., Christian de Courten, Laurence J. Garey, and Hendrik Van der Loos. 1982. “Synaptogenesis in Human Visual Cortex — Evidence for Synapse Elimination during Normal Development.” *Neuroscience Letters* 33 (3): 247–52. https://doi.org/10.1016/0304-3940(82)90379-2.

Kentro, James A., Gunjan Singh, Tuan M. Pham, Justin Currie, Saniya Khullar, Audrey T. Medeiros, Maria Tsiarli, Erica Larschan, and Kate M. O’Connor-Giles. 2025. “Conserved

Transcription Factors Coordinate Synaptic Gene Expression through Repression." *bioRxiv*, February, 2024.10.30.621128. https://doi.org/10.1101/2024.10.30.621128.

Kurmangaliyev, Yerbol Z., Juyoun Yoo, Javier Valdes-Aleman, Piero Sanfilippo, and S. Lawrence Zipursky. 2020. "Transcriptional Programs of Circuit Assembly in the Drosophila Visual System." *Neuron* 108 (6): 1045-1057.e6. https://doi.org/10.1016/j.neuron.2020.10.006.

Levine, Jacob H., Erin F. Simonds, Sean C. Bendall, Kara L. Davis, El-ad D. Amir, Michelle D. Tadmor, Oren Litvin, et al. 2015. "Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells That Correlate with Prognosis." *Cell* 162 (1): 184–97. https://doi.org/10.1016/j.cell.2015.05.047.

Marsh, Samuel, Maëlle Salmon, and Paul Hoffman. 2024. "Samuel-Marsh/scCustomize: Version 2.1.2." Zenodo. https://doi.org/10.5281/zenodo.10724532.

McInnes, Leland, John Healy, and James Melville. 2020. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." arXiv. https://doi.org/10.48550/arXiv.1802.03426.

McKeith, I., R. Perry, E. Perry, S. Lloyd, J. Court, C. Ballard, D. Spurden, N. Cairns, and R. Kerwin. 1999. "Neuronal Nicotinic Receptors in Dementia with Lewy Bodies and Schizophrenia: Alpha-Bungarotoxin and Nicotine Binding in the Thalamus." *Journal of Neurochemistry*. https://doi.org/10.1046/j.1471-4159.1999.0731590.x.

Peyré, Gabriel, and Marco Cuturi. 2019. "Computational Optimal Transport: With Applications to Data Science." *Foundations and Trends® in Machine Learning* 11 (5–6): 355–607. https://doi.org/10.1561/2200000073.

Qiu, Xiaojie, Shanshan Ding, and Tieliu Shi. 2012. "From Understanding the Development Landscape of the Canonical Fate-Switch Pair to Constructing a Dynamic Landscape for

Two-Step Neural Differentiation." *PLOS ONE* 7 (12): e49271. https://doi.org/10.1371/journal.pone.0049271.

Qiu, Xiaojie, Yan Zhang, Jorge D. Martin-Rufino, Chen Weng, Shayan Hosseinzadeh, Dian Yang, Angela N. Pogson, et al. 2022. "Mapping Transcriptomic Vector Fields of Single Cells." *Cell* 185 (4): 690-711.e45. https://doi.org/10.1016/j.cell.2021.12.045.

Schiebinger, Geoffrey, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, et al. 2019. "Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming." *Cell* 176 (4): 928-943.e22. https://doi.org/10.1016/j.cell.2019.01.006.

Sebastian, Rebecca, Kang Jin, Narciso Pavon, Ruby Bansal, Andrew Potter, Yoonjae Song, Juliana Babu, et al. 2023. "Schizophrenia-Associated NRXN1 Deletions Induce Developmental-Timing- and Cell-Type-Specific Vulnerabilities in Human Brain Organoids." *Nature Communications* 14 (June):3770. https://doi.org/10.1038/s41467-023-39420-6.

Stephens, Sarah H., Judith Logel, Amanda Barton, Alexis Franks, Jessica Schultz, Margaret Short, Jane Dickenson, et al. 2009. "Association of the 5′-Upstream Regulatory Region of the α7 Nicotinic Acetylcholine Receptor Subunit Gene (*CHRNA7*) with Schizophrenia." *Schizophrenia Research* 109 (1): 102–12. https://doi.org/10.1016/j.schres.2008.12.017.

Stuart, Tim, Avi Srivastava, Shaista Madad, Caleb A. Lareau, and Rahul Satija. 2021. "Single-Cell Chromatin State Analysis with Signac." *Nature Methods* 18 (11): 1333–41. https://doi.org/10.1038/s41592-021-01282-5.

Uzquiano, Ana, Amanda J. Kedaigle, Martina Pigoni, Bruna Paulsen, Xian Adiconis, Kwanho Kim, Tyler Faits, et al. 2022. "Proper Acquisition of Cell Class Identity in Organoids

Allows Definition of Fate Specification Programs of the Human Cerebral Cortex." *Cell* 185 (20): 3770-3788.e27. https://doi.org/10.1016/j.cell.2022.09.010.

Waltman, Ludo, and Nees Jan van Eck. 2013. "A Smart Local Moving Algorithm for Large-Scale Modularity-Based Community Detection." *The European Physical Journal B* 86 (11): 471. https://doi.org/10.1140/epjb/e2013-40829-0.

Wang, Ping, Chaoming Song, Hang Zhang, Zhanghan Wu, Xiao-Jun Tian, and Jianhua Xing. 2014. "Epigenetic State Network Approach for Describing Cell Phenotypic Transitions." *Interface Focus* 4 (3): 20130068. https://doi.org/10.1098/rsfs.2013.0068.

Xu, Chen, and Zhengchang Su. 2015. "Identification of Cell Types from Single-Cell Transcriptomes Using a Novel Clustering Method." *Bioinformatics* 31 (12): 1974–80. https://doi.org/10.1093/bioinformatics/btv088.

Zenk, Fides, Jonas Simon Fleck, Sophie Martina Johanna Jansen, Bijan Kashanian, Benedikt Eisinger, Małgorzata Santel, Jean-Samuel Dupré, J. Gray Camp, and Barbara Treutlein. 2024. "Single-Cell Epigenomic Reconstruction of Developmental Trajectories from Pluripotency in Human Neural Organoid Systems." *Nature Neuroscience* 27 (7): 1376–86. https://doi.org/10.1038/s41593-024-01652-0.

Zhang, Jiaqi, Manav Chakravarthy, and Ritambhara Singh. 2024. "scMultiNODE : Integrative Model for Multi-Modal Temporal Single-Cell Data." *bioRxiv*, October, 2024.10.27.620531. https://doi.org/10.1101/2024.10.27.620531.

Zhu, Kaiyi, Jaroslav Bendl, Samir Rahman, James M. Vicari, Claire Coleman, Tereza Clarence, Ovaun Latouche, et al. 2023. "Multi-Omic Profiling of the Developing Human Cerebral Cortex at the Single-Cell Level." *Science Advances* 9 (41): eadg3754. https://doi.org/10.1126/sciadv.adg3754.

Chapter 4: Recap and Future Directions

4.1 Summary

Development and aging are dynamic processes, involving the interaction of complex regulatory machinery at the molecular level. In many different species, males and females exhibit sexually dimorphic traits, with differential lifespan and aging-related dysfunction. Why such differences arise between biological sexes, and the mechanisms underlying such differences, continue to remain unclear—modern researchers have implemented multi-omics data integration to begin explaining these differences at the gene regulatory level. In this thesis, I investigated how the regulation of synaptic gene networks—particularly through alternative splicing and transcription—may contribute to sex-biased neurodevelopmental differences and aging trajectories.

In Chapter 2, I conducted transcriptomic analysis on *Drosophila melanogaster* third instar and adult bulk brain tissue. In doing so, I identified distinct patterns of synaptic alternative splicing (AS) events that varied by both developmental stage and biological sex. Adult males appeared to exhibit a disproportionately higher number of uniquely spliced synaptic isoforms and enriched GO pathways compared to females—suggesting that sex-biased splicing becomes more pronounced as *Drosophila* develop and age. Notably, I found sex-specific splice variants of nAChRα7 and dpr15, genes involved in a number of synaptic function and neuronal signaling pathways. Gene ontology semantic clustering analysis further revealed distinct functional categories linked to these synaptic isoforms, reinforcing the notion that alternative splicing contributes to sexually dimorphic aging in the brain.

In Chapter 3, I extended my analysis of *Drosophila* synaptic genes to human-derived cells to assess the evolutionary conservation of synaptic gene regulation. By using a set of 332 high-confidence human orthologs of *Drosophila* synaptic genes, I demonstrated temporally

coordinated expression across multiple human datasets spanning from the embryonic stage of adulthood. However, single-cell chromatin accessibility appeared inversely related to synaptic gene expression, hinting at dynamic chromatin modeling and possible TF-specific regulation across each developmental time point. Additionally, a subset of *Drosophila*-identified synaptic genes, including those involved in glutamatergic signaling, exhibited differential expression patterns in human cortex tissue that mirrors developmental transitions in excitatory neurons.

In this final chapter, I introduce the transcription factor *onecut* as a candidate regulator of synaptic genes in *D. melanogaster*, and provide both an experimental design and preliminary results to characterize its functional role in synaptic development.

4.2 Open Questions

4.2.1 AS of synaptic genes across *Drosophila* development

I observed clear sex-biased differences in the alternative splicing of synaptic genes from *Drosophila* third instar larvae to adult flies. These differences were clear not just across developmental stages, but in a sex-biased manner as well, and most prominently at the adult stage. Notably, I identified two distinct transcript isoforms of nAChα7 that appeared to be differentially upregulated in male and female adults, suggesting a sex-specific splicing mechanism involved in excitatory neuronal signaling. These differences also extended to GO analysis, as I found developmental- and sex-biased pathway enrichment across the GO sub-ontologies.

Beyond the characterization of sex-biased splicing and pathway enrichment, there are several questions that remain to be answered. Perhaps most importantly is the characterization of synaptic splicing events across the *Drosophila* developmental continuum. Does the splicing of synaptic genes change dynamically from embryogenesis through late adulthood? Akin to the

analysis conducted by Kentro et al., a future experiment should seek to incorporate several more developmental time points. In doing so, sex-biased patterns of splicing and gene expression may become more distinct, giving insight into the regulatory mechanisms that contribute to differences in synaptic development and longevity between the sexes. Similarly, I only analyzed short-read sequencing with bulk tissue—using long-read sequencing, as well as considering developmentally-coordinated synaptic expression at the single-cell resolution—will identify cellular subpopulations that seem particularly impacted by differential splicing.

A recently developed extension of *STAR* for transcriptome alignment, *STARsolo*, allows for modular, computationally-efficient mapping and quantification of scRNA-seq data (Kaminow et al., 2021). *STARsolo* successfully outperformed other commonly used pseudoalignment-to-transcriptome methods, such as *CellRanger* (Zheng et al., 2017), *kallisto-bustools* (Melsted et al., 2021) and *Alevin* (Srivastava et al., 2019), on both simulated and real-world scRNA-seq datasets. Moreover, *STARsolo* successfully detects alternative splicing, isoform expression, polyadenylation and allele-specific expression in a cell-type-specific manner (Kaminow et al., 2021). By using such a method, coupled with multi omics integration pipelines like Model-based AnalysEs of Transcriptome and RegulOme (MAESTRO) (Wang et al., 2020), it will be possible to identify candidate regulators that coordinately alter both the transcriptomic and epigenomic landscapes in a sex-biased manner.

4.2.2 Conserved synaptic gene expression and chromatin accessibility patterns in *Drosophila*, mice and humans

Moving from bulk analysis in *Drosophila* to single-cell analysis of human cells, I identified similar gene expression trends that recapitulated across different lines and recorded timepoints, and appeared to follow transcriptomic trends observed across major *Drosophila*

developmental stages. One seemingly puzzling aspect of my findings, however, was the divergence of chromatin accessibility at synaptic loci from synaptic gene expression in the Fleck et al. (2022) dataset, over the course of cell development. I expected chromatin accessibility to increase prior to synaptic gene transcription, yet observed a stark decrease in chromatin accessibility prior to peak synaptic gene expression as the human cells matured (Figure 17C).

Does this observed pattern reflect more advanced regulatory machinery—specific epigenetic and transcriptional temporal decoupling of synaptic genes—that is human specific? A prior single-cell study of *D. melanogaster* embryonic development identified the pioneer TF *Zelda* as opening chromatin regions much earlier than gene transcription in late embryonic stages (Calderon et al., 2022). Moreover, chromatin accessibility at the gene *Kah* becomes accessible in advance of gene transcription, while at the same time, the gene *Mhc* displays coordinated chromatin accessibility and gene expression (Calderon et al., 2022). Thus, chromatin accessibility does not *always* directly precede increased gene expression in *Drosophila* embryonic development, as revealed through this study, in addition to several other multi-omics analyses (Cusanovich et al., 2018; Reddington et al., 2020).

What, then, is driving this kind of temporally distinct chromatin remodeling in human cells, which are developing for a far greater period of time than *Drosophila* embryos? By identifying conserved synaptic and neuronal development-related gene activity across other species, we may be able to identify evolutionarily conserved "decoupled" gene loci—ones that are under complex regulatory control. It is also possible that the observed temporal decoupling is driven by technical artifacts in the human cerebral organoid model used—a future single-cell analysis may benefit from directly using human tissue, which will disentangle distinct biological mechanisms from model-related limitations.

Together, these findings highlight the importance of examining how upstream regulatory mechanisms dynamically influence future gene expression. While genome-wide analyses provide a clear overview of temporally coordinated gene expression and chromatin remodeling during development, identifying changes at the cellular level are equally as important—a more clear understanding of how transcription factory activity changes across development could provide a cohesive framework for better understanding gene regulation. Which transcription factors coordinate synaptic gene transcription across development and biological sexes? A compelling candidate that has emerged from systems-level and multimodal analyses is the Onecut family of TFs.

4.3 Candidate transcriptional synaptic regulation - Onecut TF family

A notable candidate transcriptional regulator of neuronal development and synaptic processes is *onecut*. Contemporary research has implicated the Onecut TF family in a myriad of developmental and neuronal processes, with conserved roles across invertebrate and mammalian species (Kropp and Gannon 2016). In mice, for example, *onecut* is required for a developmental phase of the A13 dopaminergic nucleus, as well as larger neuroanatomical structures, like the locus coeruleus and mesencephalic trigeminal nucleus (Espana and Clotman 2012). In the marine invertebrate *Ciona robusta*, *onecut* acts as a core regulator of neurotransmission in photoreceptor cells, with downstream targets conserved across numerous species for retinal cell differentiation (Vassalli et al. 2021).

In *Drosophila*, *onecut* is expressed in the brain, ventral nerve cord (VNC), sensory complexes, and stomatogastric nervous system during embryogenesis, demonstrating its importance in organ system-level developmental processes (Hammonds et al. 2013; Fisher et al. 2023). It also appears to act as a positive regulator of pan-neuronal gene transcription in *C.*

elegans, potentially interacting or competing with other transcriptional regulators like DEAF1 and CLAMP. (Leyva-Díaz and Hobert 2022; Leyva-Díaz 2023). Kentro et al. integrated scRNA-seq, scATAC-seq, and protein-protein interaction (PPI) data to predict GRNs that co-regulate synaptic gene expression and binding motifs across *Drosophila* development (Kentro et al. 2024; Khullar et al. 2024). They identified *onecut* among ten TFs that co-regulate 280 target synaptic genes across cell types and developmental time points (Kentro et al. 2024). Given its expression in the *Drosophila* central nervous system (CNS) across developmental stages, *onecut* appears to be a compelling candidate for investigating the transcriptional regulation of sex-biased synaptic aging.

4.3.1 Experimental Design - In vivo validation of *onecut*

To investigate the role of *onecut* in synaptic development, I acquired two different RNA interference (RNAi) lines, which I denote *onecut-A* and *onecut-B* (Perkins et al., 2015), and then created genetic crosses. The *onecut-A* line is encoded on chromosome 3, while *onecut-B* is encoded on chromosome 2—both lines target *onecut*, which is on chromosome 4. In my preliminary experiment, these lines were driven by *elav-Gal4*, a pan-neuronal driver, in order to ensure knockdown within the *Drosophila* CNS (Figure S8). Bouton morphology at the third instar larval neuromuscular junction (NMJ) was selected as a phenotypic readout, given their role in functional synaptic transmission.

4.3.2 Preliminary Results

At the time of writing, no clear morphological differences in synaptic bouton number were observed between control (*elav-Gal4 > GFP-RNAi*) and knockdown (*elav-Gal4 > onecut-RNAi*) lines. Several confounding variables limited interpretation. Notably, when generating the

elav-Gal4 > onecut-RNAi line, I used the *onecut*-B line, which occasionally segregated with the *Cyo* balancer. As I conducted micro dissections on third instar larvae, I was unable to determine whether or not larvae exhibited this phenotype, since they did not reach adulthood. I have since completed genetic balancing, and stable stocks have been created for both knockdown and control lines (Figure 20). Although quantification of synaptic boutons remains ongoing, I have acquired representative confocal images of larval NMJs for preliminary comparison (Figure 21).

4.3.3 Discussion & Conclusions

Predictive GRN and prior studies suggest that *onecut* may act as a primary regulator of neuronal development and maintenance. How *onecut* functionally impacts *D. melanogaster* development at the cellular level remains unclear. With my experimental crosses, it will be possible to identify whether synaptic boutons are quantitatively affected at the third instar larval NMJ. Identifying additional morphological changes induced by *onecut* overexpression or knockdown beyond bouton count—in an automated way—appears to be certainly possible. Currently, bouton counting is predominantly conducted using genotypic blinding in the image software FIJI (see 4.4.2). Prior research has developed image analysis algorithms that can be directly input as FIJI-compatible macros, allowing for the quantification of not just boutons, but nine total synaptic parameters, resulting in five distinct morphometric groupings: NMJ size, geometry, muscle size, NMJ island counting, and active zone counting (Nijhof et al., 2016; Castells-Nobau et al., 2017). Utilizing algorithmic implementations directly with imaging software, or even training image-based machine learning models for NMJ morphology classification (Minty et al., 2020; Maza et al., 2021; Singh et al., 2023), will allow for a more detailed characterization of the regulatory role of *onecut* at the molecular level.

Over the course of development, an organism experiences dynamic genetic and epigenetic remodeling—in turn, leading to phenotypic differences as it matures. Developing a system-level understanding of the regulatory mechanisms underlying these developmental processes—as well as their conservation across various species—remains a monumental task. By investigating how regulation changes across several different species—like alternative splicing, gene expression and chromatin accessibility—researchers will be able to gain insight into developmentally-coordinated activity, especially the processes that act in a sex-biased manner.

4.4 Materials and Methods

4.4.1 *Drosophila* stocks

The following fly lines used in this study were obtained from the Bloomington Drosophila Stock Center (BDSC): *elav*^{C155}*-Gal4* (RRID:BDSC_458), *UAS-GFP-RNAi* (RRID:BDSC_9330), *UAS-onecut*^{HMS01438} (RRID:BDSC_35025), and *UAS-onecut*^{HMJ22466} (RRID:BDSC_58336). Balancer fly lines maintained in our lab (used to generate the correct crosses) are as follows: double balancer (*w*DD/*w*DD; *IF*/*CyO*; *MKRS*/*TM6Tb*), single balancer (*w*DD/*w*DD;; *MKRS*/*TM6Tb*), single balancer (*w*DD/*w*DD; *IF*/*CyO*). All *Drosophila melanogaster* crosses were set up with 10 males and 10 females and raised on molasses food (Lab Express, Type R) in a 25°C incubator with controlled humidity and 12h light/dark cycle.

4.4.2 Phenotypic analysis of synapse number

Male third-instar larvae were sexed and dissected (Figure S9) in ice-cold Ca2+-free saline and fixed for 6 minutes in Bouin's Fixative (Electron Microscopy Sciences, Cat# 15990). Dissections were washed several times with 1X PBS and then permeabilized with 1X PBS containing 0.1% Triton-X (Fisher Scientific, Cat# BP151-100). Dissected larvae were then blocked either for 1

hour at room temperature or overnight at 4°C in 1X PBS containing 0.1% Triton-X and 1% BSA (Sigma-Aldrich, Cat# A7906). Dissections were incubated in primary antibodies overnight at 4°C and in secondary antibodies for 2–4 hours at room temperature or overnight at 4°C. Dissected larvae were mounted in Vectashield (Vector Laboratories, Cat# H-1000-10). The following antibodies were used at the indicated concentrations: mouse anti-Bruchpilot (Brp) at 1:100 (DSHB, Cat# nc82, RRID: AB_2314866), Alexa Fluor 488 goat anti-mouse at 1:500 (Invitrogen, Cat# A-11029, RRID: AB_2534088) and anti-Horseradish Peroxidase (HRP) conjugated to Alexa Fluor 647 at 1:500 (Jackson ImmunoResearch Labs, Cat# 123-605-021, RRID: AB_2338967).

Larval NMJ images were taken on a Nikon A1R HD confocal microscope with a 40x oil immersion objective. Boutons were quantified at the highly stereotyped larval NMJ4 muscle segments, A2-A4. A bouton was defined as a bulbous swelling highlighted by the neuronal membrane marker HRP and co-stained with the active zone marker Brp. Quantification of bouton number was performed masked to genotype/condition using the “multi-point” tool in FIJI/ImageJ (NIH, RRID: SCR_003070). Sample sizes were based on a prior published study.

4.5 Figures

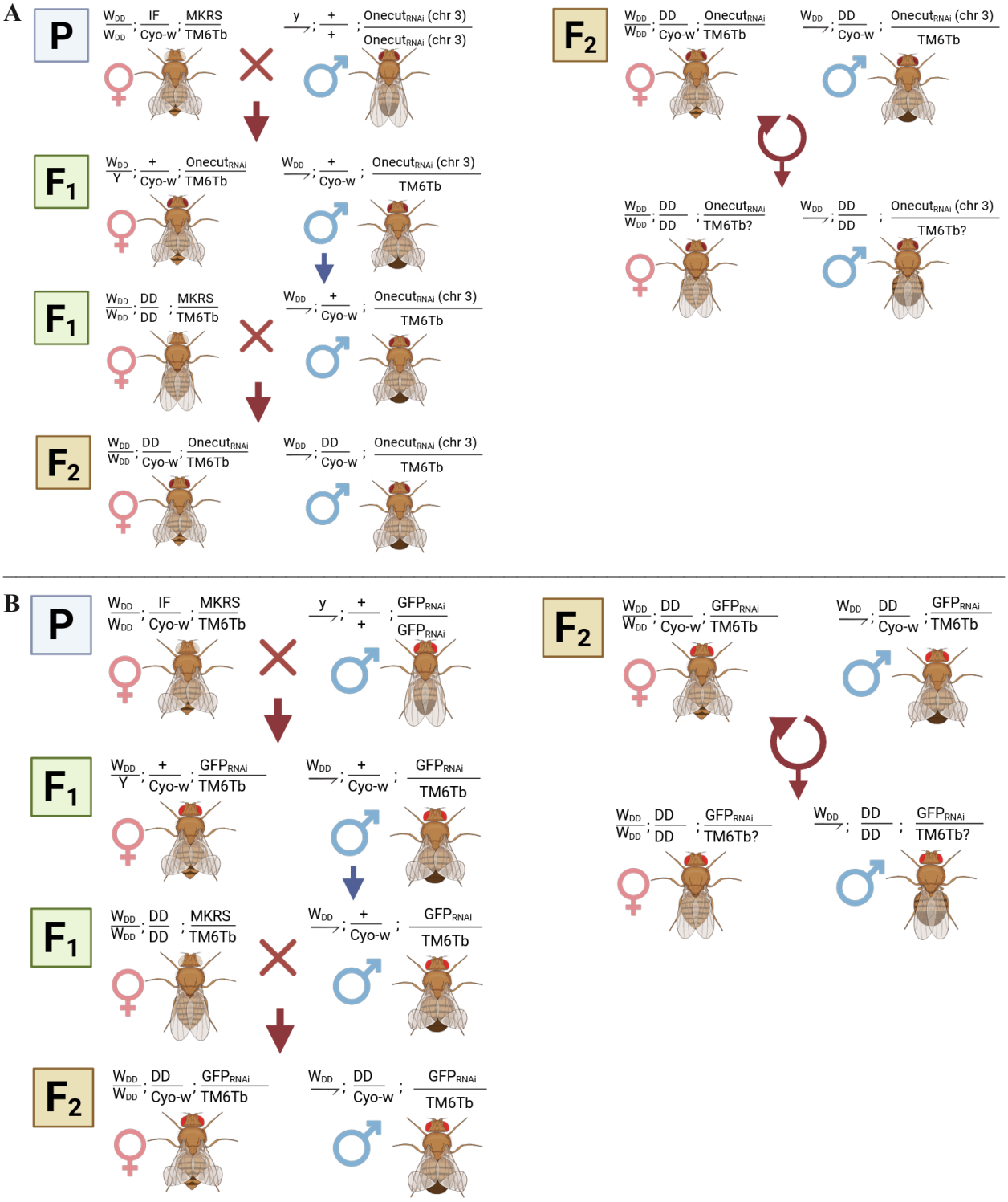


Figure 20: Crossing schemes for *onecut-A* and control lines. (A) Experimental (*onecut-RNAi A line*) crosses to create viable stock for larval bouton quantification. **(B)** Control (*GFP-RNAi*) crosses to create a viable comparison for the experimental *onecut-A* line. Created with BioRender.com.

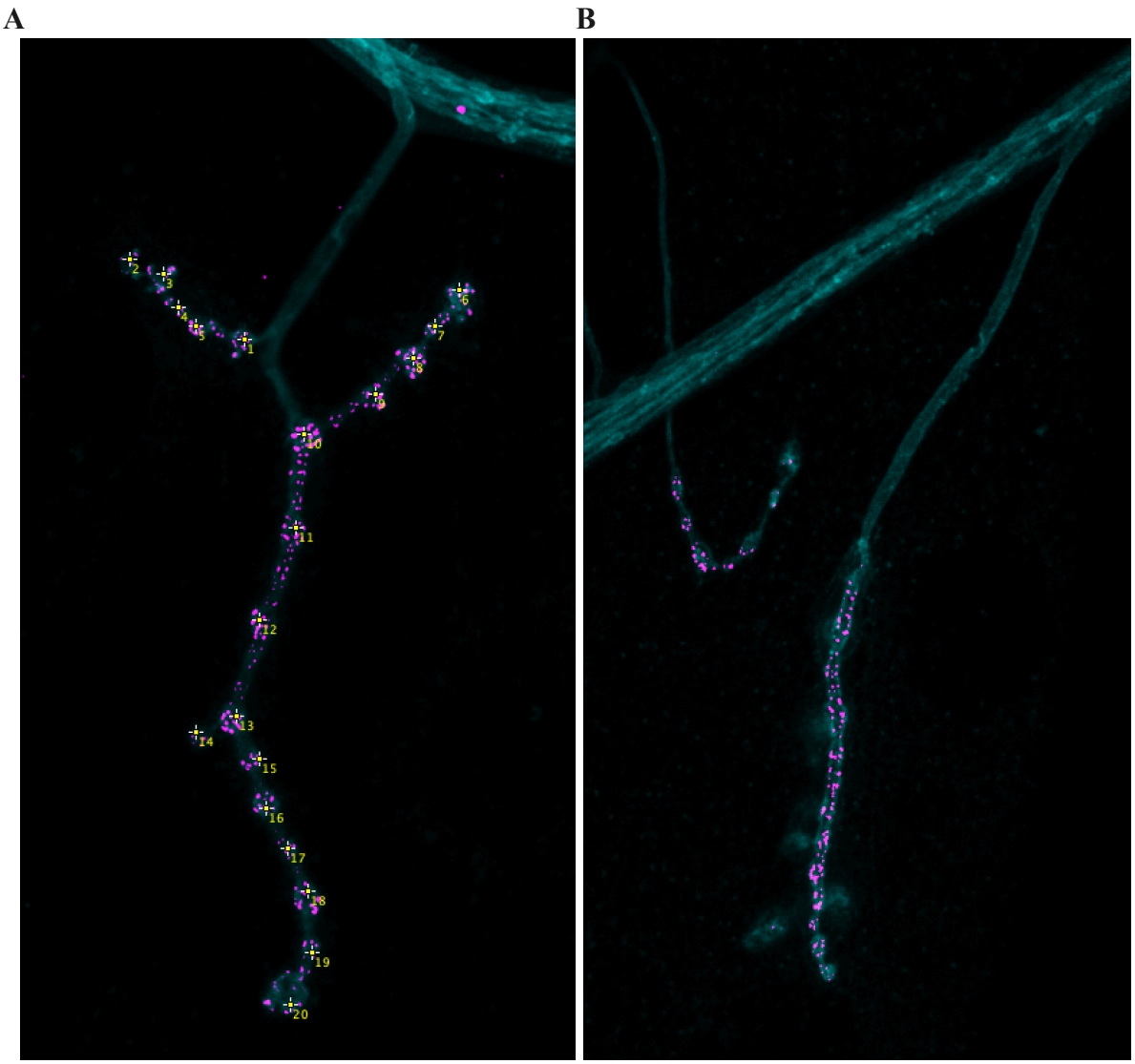


Figure 21: Representative confocal images of *Drosophila melanogaster* third instar larval NMJ A4. High-magnification maximum intensity projections of two confocal images of larval NMJ A4 at 75% image-size. **(A)** NMJ of male elav-Gal4 > $Onecut_{RNAi}$B third instar larva. Multipoint tool labels represented by yellow and white plus-shaped marks. **(B)** NMJ of male elav-Gal4 > GFP_{RNAi} third instar larva. Synaptic boutons are highlighted by the neuronal membrane marker HRP in pink, and co-stained with the active zone marker Brp in blue.

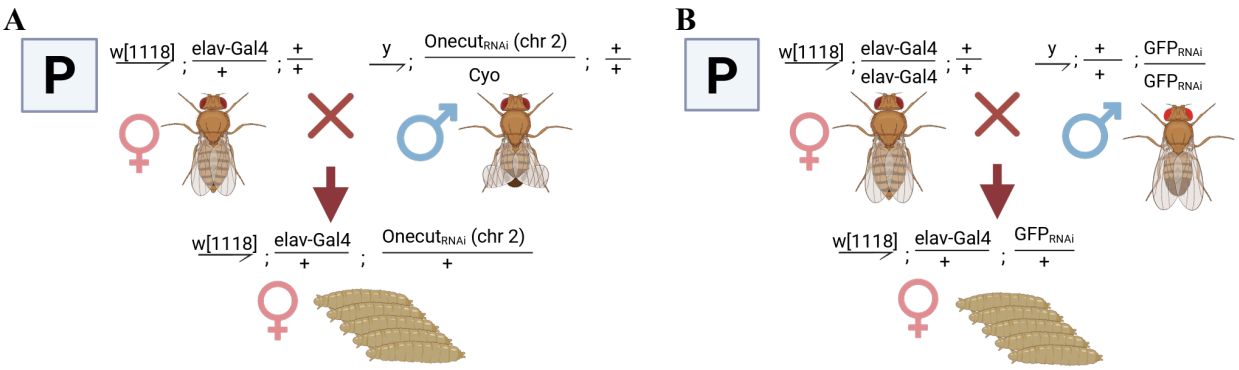


Figure S8: Crossing schemes for *onecut-B* and control lines. (A) Experimental (*onecut-RNAi-B line*) and **(B)** control (*GFP-RNAi*) crosses using *elav-Gal4* driver. Created with BioRender.com.

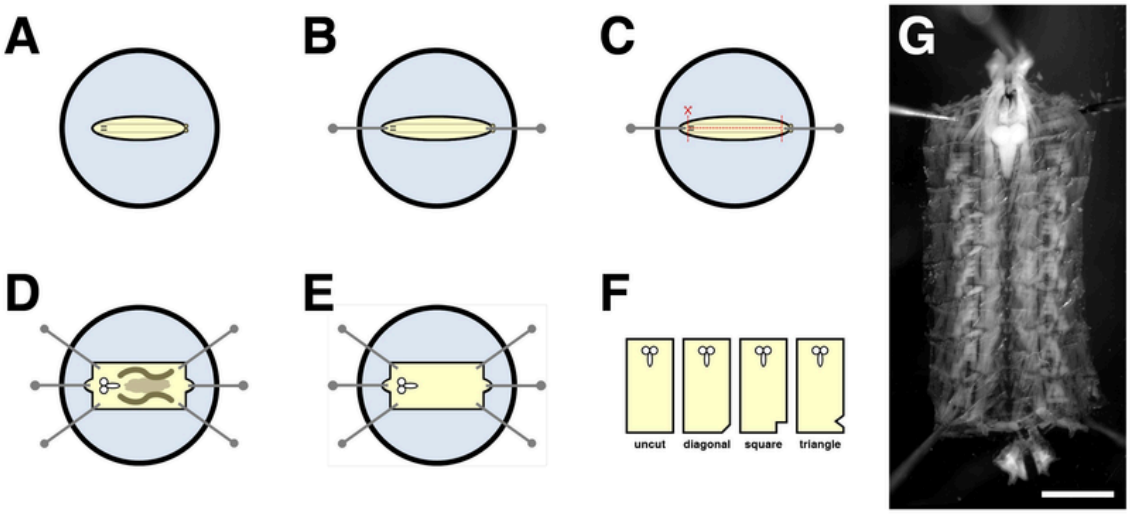


Figure S9: Third instar larvae microdissection steps. (A) A larva is placed on a sylgard disc, dorsal side up. **(B)** Larva pinned at anterior and posterior ends with pins, fully stretched out. **(C)** Horizontal cuts are made with microdissection scissors on each end of the larva. **(D)** The body wall is unfurled and pinned down at each corner to resemble a rectangle shape, followed by **(E)** cleaning of the internal organs. **(F)** Methods of distinguishing genotypes after dissection. **(G)** Stretched and pinned body wall, scale bar = 1mm. Adapted from Wang et al. (2015).

4.6 References

Calderon, Diego, Ronnie Blecher-Gonen, Xingfan Huang, Stefano Secchia, James Kentro, Riza M. Daza, Beth Martin, et al. 2022. "The Continuum of Drosophila Embryonic Development at Single-Cell Resolution." *Science (New York, N.Y.)* 377 (6606): eabn5800. https://doi.org/10.1126/science.abn5800.

Castells-Nobau, Anna, Bonnie Nijhof, Ilse Eidhof, Louis Wolf, Jolanda M. Scheffer-de Gooyert, Ignacio Monedero, Laura Torroja, Jeroen A.W.M. van der Laak, and Annette Schenck. 2017. "Two Algorithms for High-Throughput and Multi-Parametric Quantification of Drosophila Neuromuscular Junction Morphology." *Journal of Visualized Experiments : JoVE*, no. 123 (May), 55395. https://doi.org/10.3791/55395.

Cusanovich, Darren A., James P. Reddington, David A. Garfield, Riza M. Daza, Delasa Aghamirzaie, Raquel Marco-Ferreres, Hannah A. Pliner, et al. 2018. "The Cis-Regulatory Dynamics of Embryonic Development at Single-Cell Resolution." *Nature* 555 (7697): 538–42. https://doi.org/10.1038/nature25981.

Espana, A., and F. Clotman. 2012. "Onecut Factors Control Development of the Locus Coeruleus and of the Mesencephalic Trigeminal Nucleus." *Molecular and Cellular Neurosciences* 50 (1): 93–102. https://doi.org/10.1016/j.mcn.2012.04.002.

Fisher, William W, Ann S Hammonds, Richard Weiszmann, Benjamin W Booth, Louis Gevirtzman, Jaeda E J Patton, Connor A Kubo, Robert H Waterston, and Susan E Celniker. 2023. "A modERN Resource: Identification of Drosophila Transcription Factor Candidate Target Genes Using RNAi." *Genetics* 223 (4): iyad004. https://doi.org/10.1093/genetics/iyad004.

Hammonds, Ann S., Christopher A. Bristow, William W. Fisher, Richard Weiszmann, Siqi Wu, Volker Hartenstein, Manolis Kellis, Bin Yu, Erwin Frise, and Susan E. Celniker. 2013.

"Spatial Expression of Transcription Factors in Drosophila Embryonic Organ Development." *Genome Biology* 14 (12): R140. https://doi.org/10.1186/gb-2013-14-12-r140.

Kaminow, Benjamin, Dinar Yunusov, and Alexander Dobin. 2021. "STARsolo: Accurate, Fast and Versatile Mapping/Quantification of Single-Cell and Single-Nucleus RNA-Seq Data." bioRxiv. https://doi.org/10.1101/2021.05.05.442755.

Kentro, James A., Gunjan Singh, Tuan M. Pham, Justin Currie, Saniya Khullar, Audrey T. Medeiros, Maria Tsiarli, Erica Larschan, and Kate M. O'Connor-Giles. 2025. "Conserved Transcription Factors Coordinate Synaptic Gene Expression through Repression." *bioRxiv*, February, 2024.10.30.621128. https://doi.org/10.1101/2024.10.30.621128.

Khullar, Saniya, Xiang Huang, Raghu Ramesh, John Svaren, and Daifeng Wang. 2024. "NetREm: Network Regression Embeddings Reveal Cell-Type Transcription Factor Coordination for Gene Regulation." *bioRxiv*, May, 2023.10.25.563769. https://doi.org/10.1101/2023.10.25.563769.

Kropp, Peter A., and Maureen Gannon. 2016. "Onecut Transcription Factors in Development and Disease." *Trends in Developmental Biology* 9:43–57.

Leyva-Díaz, Eduardo. 2023. "CUT Homeobox Genes: Transcriptional Regulation of Neuronal Specification and Beyond." *Frontiers in Cellular Neuroscience* 17 (September):1233830. https://doi.org/10.3389/fncel.2023.1233830.

Leyva-Díaz, Eduardo, and Oliver Hobert. 2022. "Robust Regulatory Architecture of Pan-Neuronal Gene Expression." *Current Biology : CB* 32 (8): 1715-1727.e8. https://doi.org/10.1016/j.cub.2022.02.040.

Mejia Maza, Alan, Seth Jarvis, Weaverly Colleen Lee, Thomas J. Cunningham, Giampietro Schiavo, Maria Secrier, Pietro Fratta, James N. Sleigh, Elizabeth M. C. Fisher, and Carole

H. Sudre. 2021. “NMJ-Analyser Identifies Subtle Early Changes in Mouse Models of Neuromuscular Disease.” *Scientific Reports* 11 (1): 12251. https://doi.org/10.1038/s41598-021-91094-6.

Melsted, Páll, A. Sina Booeshaghi, Lauren Liu, Fan Gao, Lambda Lu, Kyung Hoi (Joseph) Min, Eduardo da Veiga Beltrame, Kristján Eldjárn Hjörleifsson, Jase Gehring, and Lior Pachter. 2021. “Modular, Efficient and Constant-Memory Single-Cell RNA-Seq Preprocessing.” *Nature Biotechnology* 39 (7): 813–18. https://doi.org/10.1038/s41587-021-00870-2.

Minty, Gavin, Alex Hoppen, Ines Boehm, Abrar Alhindi, Larissa Gibb, Ellie Potter, Boris C. Wagner, et al. 2020. “aNMJ-Morph: A Simple Macro for Rapid Analysis of Neuromuscular Junction Morphology.” *Royal Society Open Science* 7 (4): 200128. https://doi.org/10.1098/rsos.200128.

Nijhof, Bonnie, Anna Castells-Nobau, Louis Wolf, Jolanda M. Scheffer-de Gooyert, Ignacio Monedero, Laura Torroja, Lluis Coromina, Jeroen A. W. M. van der Laak, and Annette Schenck. 2016. “A New Fiji-Based Algorithm That Systematically Quantifies Nine Synaptic Parameters Provides Insights into Drosophila NMJ Morphometry.” *PLOS Computational Biology* 12 (3): e1004823. https://doi.org/10.1371/journal.pcbi.1004823.

Perkins, Lizabeth A, Laura Holderbaum, Rong Tao, Yanhui Hu, Richelle Sopko, Kim McCall, Donghui Yang-Zhou, et al. 2015. “The Transgenic RNAi Project at Harvard Medical School: Resources and Validation.” *Genetics* 201 (3): 843–52. https://doi.org/10.1534/genetics.115.180208.

Reddington, James P., David A. Garfield, Olga M. Sigalova, Aslihan Karabacak Calviello, Raquel Marco-Ferreres, Charles Girardot, Rebecca R. Viales, Jacob F. Degner, Uwe Ohler, and Eileen E. M. Furlong. 2020. “Lineage-Resolved Enhancer and Promoter Usage during a

Time Course of Embryogenesis." *Developmental Cell* 55 (5): 648-664.e9. https://doi.org/10.1016/j.devcel.2020.10.009.

Singh, Jaskaran, Yingzhou Edward Pan, and Shunmoogum A Patten. 2023. "NMJ Analyser: A Novel Method to Quantify Neuromuscular Junction Morphology in Zebrafish." *Bioinformatics* 39 (12): btad720. https://doi.org/10.1093/bioinformatics/btad720.

Srivastava, Avi, Laraib Malik, Tom Smith, Ian Sudbery, and Rob Patro. 2019. "Alevin Efficiently Estimates Accurate Gene Abundances from dscRNA-Seq Data." *Genome Biology* 20 (1): 65. https://doi.org/10.1186/s13059-019-1670-y.

Vassalli, Quirino Attilio, Chiara Colantuono, Valeria Nittoli, Anna Ferraioli, Giulia Fasano, Federica Berruto, Maria Luisa Chiusano, Robert Neil Kelsh, Paolo Sordino, and Annamaria Locascio. 2021. "Onecut Regulates Core Components of the Molecular Machinery for Neurotransmission in Photoreceptor Differentiation." *Frontiers in Cell and Developmental Biology* 9:602450. https://doi.org/10.3389/fcell.2021.602450.

Wang, Chenfei, Dongqing Sun, Xin Huang, Changxin Wan, Ziyi Li, Ya Han, Qian Qin, et al. 2020. "Integrative Analyses of Single-Cell Transcriptome and Regulome Using MAESTRO." *Genome Biology* 21 (1): 198. https://doi.org/10.1186/s13059-020-02116-x.

Wang, Simon, SooHyun Yoo, Hae-yoon Kim, Mannan Wang, Clare Zheng, Wade Parkhouse, Charles Krieger, and Nicholas Harden. 2015. "Detection of In Situ Protein-Protein Complexes at the Drosophila Larval Neuromuscular Junction Using Proximity Ligation Assay." *Journal of Visualized Experiments : JoVE*, no. 95 (January), 52139. https://doi.org/10.3791/52139.

Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. "Massively Parallel Digital Transcriptional

Profiling of Single Cells." *Nature Communications* 8 (1): 14049.

https://doi.org/10.1038/ncomms14049.

Appendix A: Top 332 human synaptic gene orthologs

FlyBaseID	Fly Symbol	Human Symbol	Ensembl ID
FBgn0038693	unc79	UNC79	ENSG00000133958
FBgn0037754	CG8500	DIRAS2	ENSG00000165023
FBgn0261041	stj	CACNA2D3	ENSG00000157445
FBgn0013973	Gycbeta100B	GUCY1B1	ENSG00000061918
FBgn0039420	CG6154	DPEP1	ENSG00000015413
FBgn0023179	amon	PCSK2	ENSG00000125851
FBgn0033911	VGAT	SLC32A1	ENSG00000101438
FBgn0036319	Ent3	SLC29A4	ENSG00000164638
FBgn0010329	Tbh	DBH	ENSG00000123454
FBgn0033932	Dh44-R1	CRHR1	ENSG00000120088
FBgn0039536	unc80	UNC80	ENSG00000144406
FBgn0000303	ChAT	CHAT	ENSG00000070748
FBgn0003091	Pkc53E	PRKCA	ENSG00000154229
FBgn0260943	Rbp6	MSI2	ENSG00000153944
FBgn0031081	Nep3	ECE1	ENSG00000117298
FBgn0261794	kcc	SLC12A4	ENSG00000124067
FBgn0033368	CG13743	SLC38A11	ENSG00000169507
FBgn0270928	VAChT	SLC18A3	ENSG00000187714
FBgn0034135	Syn2	SNTG1	ENSG00000147481
FBgn0266377	Pde8	PDE8B	ENSG00000113231
FBgn0030613	Rab3-GEF	MADD	ENSG00000110514
FBgn0030706	Lrp4	LRP4	ENSG00000134569
FBgn0005564	Shal	KCND2	ENSG00000184408
FBgn0003093	Pkc98E	PRKCE	ENSG00000171132
FBgn0264386	Ca-alpha1T	CACNA1I	ENSG00000100346
FBgn0085390	Dgk	DGKB	ENSG00000136267
FBgn0034136	DAT	SLC6A2	ENSG00000103546
FBgn0035649	CG10483	XPR1	ENSG00000143324

FBgn0039915	Gat	SLC6A1	ENSG00000157103
FBgn0028481	CG4341	TMTC2	ENSG00000179104
FBgn0259221	ATP8A	ATP8A1	ENSG00000124406
FBgn0260446	GABA-B-R1	GABBR1	ENSG00000204681
FBgn0010014	CanB	PPP3R1	ENSG00000221823
FBgn0033500	CG12913	CSGALNACT1	ENSG00000147408
FBgn0263456	nwk	FCHSD2	ENSG00000137478
FBgn0039770	CG15537	CRHBP	ENSG00000145708
FBgn0033876	Syngr	SYNGR1	ENSG00000100321
FBgn0031765	CG9109	B3GLCT	ENSG00000187676
FBgn0040348	CG3703	RUNDC1	ENSG00000198863
FBgn0034567	CG15651	FKRP	ENSG00000181027
FBgn0004516	Gad1	GAD1	ENSG00000128683
FBgn0033054	CG14591	TMEM164	ENSG00000157600
FBgn0022382	Pka-R2	PRKAR2A	ENSG00000114302
FBgn0033031	CG8245	TMEM53	ENSG00000126106
FBgn0035245	GC	GGCX	ENSG00000115486
FBgn0032593	Trpgamma	TRPC4	ENSG00000133107
FBgn0033500	CG12913	CSGALNACT2	ENSG00000169826
FBgn0028704	Nckx30C	SLC24A2	ENSG00000155886
FBgn0032013	Scgalpha	SGCE	ENSG00000127990
FBgn0034789	PIP5K59B	PIP5K1A	ENSG00000143398
FBgn0039911	CG1909	RAPSN	ENSG00000165917
FBgn0031716	CG14015	MANEA	ENSG00000172469
FBgn0265262	Vha68-1	ATP6V1A	ENSG00000114573
FBgn0000346	comt	NSF	ENSG00000073969
FBgn0035187	Trh	TPH1	ENSG00000129167
FBgn0031424	VGlut	SLC17A6	ENSG00000091664
FBgn0002917	na	NALCN	ENSG00000102452
FBgn0003386	Shaw	KCNC2	ENSG00000166006

FBgn0010240	Lcch3	GABRB3	ENSG00000166206
FBgn0024941	RSG7	RGS7	ENSG00000182901
FBgn0029105	alpha-Catr	CTNNAL1	ENSG00000119326
FBgn0019985	mGluR	GRM3	ENSG00000198822
FBgn0052944	CG32944	STK32A	ENSG00000169302
FBgn0052944	CG32944	STK32B	ENSG00000152953
FBgn0003386	Shaw	KCNC1	ENSG00000129159
FBgn0032151	nAChRalpha6	CHRNA7	ENSG00000175344
FBgn0032151	nAChRalpha6	CHRFAM7A	ENSG00000166664
FBgn0086778	nAChRalpha7	CHRNA7	ENSG00000175344
FBgn0086778	nAChRalpha7	CHRFAM7A	ENSG00000166664
FBgn0036544	sff	BRSK1	ENSG00000160469
FBgn0028875	nAChRalpha5	CHRNA7	ENSG00000175344
FBgn0030897	Frq1	NCS1	ENSG00000107130
FBgn0035733	CG8641	RASD1	ENSG00000108551
FBgn0083228	Frq2	NCS1	ENSG00000107130
FBgn0038641	ChT	SLC5A7	ENSG00000115665
FBgn0052843	Dh31-R	CALCRL	ENSG00000064989
FBgn0000108	Appl	APLP2	ENSG00000084234
FBgn0038165	Task6	KCNK3	ENSG00000171303
FBgn0038165	Task6	KCNK15	ENSG00000124249
FBgn0030358	CG10362	PDZD8	ENSG00000165650
FBgn0037690	Task7	KCNK3	ENSG00000171303
FBgn0037690	Task7	KCNK15	ENSG00000124249
FBgn0260964	Vmat	SLC18A1	ENSG00000036565
FBgn0023416	Ac3	ADCY3	ENSG00000138031
FBgn0027575	GABA-B-R2	GABBR2	ENSG00000136928
FBgn0004619	GluRIA	GRIA3	ENSG00000125675
FBgn0010399	Nmdar1	GRIN1	ENSG00000176884
FBgn0085413	CG34384	DGKH	ENSG00000102780

FBgn0003429	slo	KCNMA1	ENSG00000156113
FBgn0267002	unc-104	KIF1A	ENSG00000130294
FBgn0035756	unc-13-4A	BAIAP3	ENSG00000007516
FBgn0013334	Sap47	SYAP1	ENSG00000169895
FBgn0025391	Scgdelta	SGCD	ENSG00000170624
FBgn0037690	Task7	KCNK9	ENSG00000169427
FBgn0033958	jef	MFSD6	ENSG00000151690
FBgn0036544	sff	BRSK2	ENSG00000174672
FBgn0053513	Nmdar2	GRIN2D	ENSG00000105464
FBgn0031064	CG12531	SLC7A14	ENSG00000013293
FBgn0039530	Tusp	TULP4	ENSG00000130338
FBgn0004622	TkR99D	TACR3	ENSG00000169836
FBgn0029834	CG5937	ARFGEF3	ENSG00000112379
FBgn0051140	CG31140	DGKQ	ENSG00000145214
FBgn0024963	GluClalpha	GLRA1	ENSG00000145888
FBgn0267252	Ggamma30A	GNG13	ENSG00000127588
FBgn0037130	Syn1	SNTB1	ENSG00000172164
FBgn0000037	mAChR-A	CHRM1	ENSG00000168539
FBgn0003380	Sh	KCNA3	ENSG00000177272
FBgn0285944	para	SCN2A	ENSG00000136531
FBgn0038839	Ktl	KCTD12	ENSG00000178695
FBgn0037130	Syn1	SNTB2	ENSG00000168807
FBgn0265959	rdgC	PPEF1	ENSG00000086717
FBgn0004242	Syt1	SYT1	ENSG00000067715
FBgn0038975	Nrx-1	NRXN1	ENSG00000179915
FBgn0000037	mAChR-A	CHRM3	ENSG00000133019
FBgn0005586	Rab3	RAB3C	ENSG00000152932
FBgn0036984	CG13248	SLC7A4	ENSG00000099960
FBgn0013972	Gycalpha99B	GUCY1A2	ENSG00000152402
FBgn0003380	Sh	KCNA1	ENSG00000111262

FBgn0285944	para	SCN8A	ENSG00000196876
FBgn0004118	nAChRbeta2	CHRNA3	ENSG00000080644
FBgn0037326	CG14669	RASL10B	ENSG00000270885
FBgn0042696	NfI	NFIC	ENSG00000141905
FBgn0259111	Ndae1	SLC4A8	ENSG00000050438
FBgn0034720	Liprin-gamma	KAZN	ENSG00000189337
FBgn0038839	Ktl	KCTD16	ENSG00000183775
FBgn0031294	IA-2	PTPRN	ENSG00000054356
FBgn0004575	Syn	SYN3	ENSG00000185666
FBgn0031065	CG14234	TMEM198	ENSG00000188760
FBgn0038042	Scgbeta	SGCB	ENSG00000163069
FBgn0004580	Cbp53E	CALB2	ENSG00000172137
FBgn0042696	NfI	NFIA	ENSG00000162599
FBgn0262483	Rbp	RIMBP2	ENSG00000060709
FBgn0034950	Pask	PASK	ENSG00000115687
FBgn0037989	ATP8B	ATP8B2	ENSG00000143515
FBgn0052333	CG32333	FAM135A	ENSG00000082269
FBgn0086913	Rab26	RAB26	ENSG00000167964
FBgn0263131	CG43373	ADCY5	ENSG00000173175
FBgn0031116	CG1695	SGSM2	ENSG00000141258
FBgn0260657	CG42540	STOM	ENSG00000148175
FBgn0052085	CG32085	FBXL16	ENSG00000127585
FBgn0039282	Bili	FRMD8	ENSG00000126391
FBgn0027491	Cdk5alpha	CDK5R1	ENSG00000176749
FBgn0264000	GluRIB	GRIA4	ENSG00000152578
FBgn0029761	SK	KCNN3	ENSG00000143603
FBgn0032946	nrv3	ATP1B1	ENSG00000143153
FBgn0011676	Nos	NOS3	ENSG00000164867
FBgn0033058	CCHa2-R	BRS3	ENSG00000102239
FBgn0001991	Ca-alpha1D	CACNA1D	ENSG00000157388

FBgn0001991	Ca-alpha1D	CACNA1S	ENSG00000081248
FBgn0011676	Nos	NOS1	ENSG00000089250
FBgn0000036	nAChRalpha1	CHRNA6	ENSG00000147434
FBgn0039380	CG5890	KCNIP1	ENSG00000182132
FBgn0083963	Nlg3	NLGN4X	ENSG00000146938
FBgn0083963	Nlg3	NLGN1	ENSG00000169760
FBgn0083963	Nlg3	NLGN3	ENSG00000196338
FBgn0263111	cac	CACNA1A	ENSG00000141837
FBgn0020269	mspo	SPON2	ENSG00000159674
FBgn0259099	DCX-EMAP	EML4	ENSG00000143924
FBgn0259099	DCX-EMAP	EML1	ENSG00000066629
FBgn0035050	SiaT	ST6GAL2	ENSG00000144057
FBgn0027654	jdp	DNAJC12	ENSG00000108176
FBgn0261804	CG42750	DPEP2	ENSG00000167261
FBgn0053147	Hs3st-A	HS3ST5	ENSG00000249853
FBgn0050418	nord	NDNF	ENSG00000173376
FBgn0028879	CG15270	ANO8	ENSG00000074855
FBgn0031835	CG11319	DPP10	ENSG00000175497
FBgn0034985	CG3328	MYRF	ENSG00000124920
FBgn0025625	Sik2	SIK2	ENSG00000170145
FBgn0039054	Cow	SPOCK3	ENSG00000196104
FBgn0041707	7B2	SCG5	ENSG00000166922
FBgn0265487	mbl	MBNL2	ENSG00000139793
FBgn0259099	DCX-EMAP	EML2	ENSG00000125746
FBgn0039054	Cow	SPOCK2	ENSG00000107742
FBgn0264815	Pde1c	PDE1A	ENSG00000115252
FBgn0266098	rg	NBEA	ENSG00000172915
FBgn0010414	SerT	SLC6A4	ENSG00000108576
FBgn0053547	Rim	RIMS1	ENSG00000079841
FBgn0053547	Rim	RIMS2	ENSG00000176406

FBgn0028400	Syt4	SYT11	ENSG00000132718
FBgn0028400	Syt4	SYT4	ENSG00000132872
FBgn0264006	dysc	WHRN	ENSG00000095397
FBgn0259241	CG42339	SBSPON	ENSG00000164764
FBgn0031299	CG4629	NIM1K	ENSG00000177453
FBgn0051619	nolo	ADAMTSL1	ENSG00000178031
FBgn0051619	nolo	ADAMTSL3	ENSG00000156218
FBgn0013342	nSyb	VAMP2	ENSG00000220205
FBgn0039396	CCAP-R	NPSR1	ENSG00000187258
FBgn0000039	nAChRalpha2	CHRNA2	ENSG00000120903
FBgn0028743	Dhit	RGS17	ENSG00000091844
FBgn0004168	5-HT1A	HTR1A	ENSG00000178394
FBgn0259110	mmd	ADAM11	ENSG00000073670
FBgn0085386	CG34357	GUCY2D	ENSG00000132518
FBgn0051646	DIP-theta	OPCML	ENSG00000183715
FBgn0029837	Tsp5D	TSPAN9	ENSG00000011105
FBgn0041605	cpx	CPLX1	ENSG00000168993
FBgn0085447	sif	TIAM1	ENSG00000156299
FBgn0000024	Ace	BCHE	ENSG00000114200
FBgn0028743	Dhit	RGS20	ENSG00000147509
FBgn0005659	Ets98B	SPDEF	ENSG00000124664
FBgn0259222	CG42322	SLC35F3	ENSG00000183780
FBgn0004513	Mdr65	ABCB4	ENSG00000005471
FBgn0034070	SP2353	EGFLAM	ENSG00000164318
FBgn0040281	Aplip1	MAPK8IP1	ENSG00000121653
FBgn0041723	rho-5	RHBDF1	ENSG00000007384
FBgn0261086	Syt14	SYT16	ENSG00000139973
FBgn0261086	Syt14	SYT14	ENSG00000143469
FBgn0024913	Actbeta	INHBA	ENSG00000122641
FBgn0037698	CG16779	ZNF541	ENSG00000118156

FBgn0259822	Ca-beta	CACNB1	ENSG00000067191
FBgn0259822	Ca-beta	CACNB2	ENSG00000165995
FBgn0024913	Actbeta	INHBB	ENSG00000163083
FBgn0262593	Shab	KCNB1	ENSG00000158445
FBgn0039883	RhoGAP100F	SYDE2	ENSG00000097096
FBgn0039883	RhoGAP100F	SYDE1	ENSG00000105137
FBgn0030634	CG9164	WSCD2	ENSG00000075035
FBgn0260660	Mp	COL15A1	ENSG00000204291
FBgn0036934	sNPF-R	PRLHR	ENSG00000119973
FBgn0283680	IP3K2	ITPKB	ENSG00000143772
FBgn0259821	CG42402	EVA1C	ENSG00000166979
FBgn0283680	IP3K2	ITPKA	ENSG00000137825
FBgn0029663	CG10804	SLC6A18	ENSG00000164363
FBgn0262562	CG43102	ARHGEF17	ENSG00000110237
FBgn0016975	stnB	STON2	ENSG00000140022
FBgn0052082	IRSp53	BAIAP2	ENSG00000175866
FBgn0264574	Glut1	SLC2A3	ENSG00000059804
FBgn0259150	OtopLc	OTOP2	ENSG00000183034
FBgn0038237	Pde6	PDE11A	ENSG00000128655
FBgn0263029	CG43324	GNG12	ENSG00000172380
FBgn0261085	Syt12	SYT12	ENSG00000173227
FBgn0064123	stg1	CACNG7	ENSG00000105605
FBgn0038653	Octalpha2R	ADRA2A	ENSG00000150594
FBgn0031414	eys	EYS	ENSG00000188107
FBgn0259246	brp	ERC1	ENSG00000082805
FBgn0053517	Dop2R	DRD2	ENSG00000149295
FBgn0259994	OtopLa	OTOP1	ENSG00000163982
FBgn0263772	CG43689	MYT1L	ENSG00000186487
FBgn0263772	CG43689	ST18	ENSG00000147488
FBgn0263772	CG43689	MYT1	ENSG00000196132

FBgn0259108	futsch	MAP1B	ENSG00000131711
FBgn0259246	brp	ERC2	ENSG00000187672
FBgn0038880	SIFaR	NPFFR2	ENSG00000056291
FBgn0267796	Tmc	TMC3	ENSG00000188869
FBgn0259935	CG42458	HNRNPCL1	ENSG00000179172
FBgn0259935	CG42458	HNRNPC	ENSG00000092199
FBgn0038880	SIFaR	NPFFR1	ENSG00000148734
FBgn0261788	Ank2	ANK2	ENSG00000145362
FBgn0034692	CG13502	ODAD4	ENSG00000204815
FBgn0261928	CG42795	TBC1D30	ENSG00000111490
FBgn0259171	Pde9	PDE9A	ENSG00000160191
FBgn0086708	stv	BAG3	ENSG00000151929
FBgn0086708	stv	BAG4	ENSG00000156735
FBgn0259146	fid	TRMT9B	ENSG00000250305
FBgn0265416	Neto	NETO2	ENSG00000171208
FBgn0085419	Rgk2	REM1	ENSG00000088320
FBgn0264542	hwt	SHF	ENSG00000138606
FBgn0004573	5-HT7	HTR7	ENSG00000148680
FBgn0265182	CG44247	NPDC1	ENSG00000107281
FBgn0038980	Octbeta1R	HTR4	ENSG00000164270
FBgn0087011	CG41520	ANGPT1	ENSG00000154188
FBgn0259927	CG42450	RGS11	ENSG00000076344
FBgn0050361	mtt	GRM8	ENSG00000179603
FBgn0052183	Ccn	CCN2	ENSG00000118523
FBgn0264339	CG43795	GPR158	ENSG00000151025
FBgn0051665	wry	DNER	ENSG00000187957
FBgn0264753	Rgk1	RRAD	ENSG00000166592
FBgn0038246	CG14853	ERICH2	ENSG00000204334
FBgn0033504	CAP	SORBS1	ENSG00000095637
FBgn0039084	CG10175	CES2	ENSG00000172831

FBgn0039084	CG10175	CES5A	ENSG00000159398
FBgn0052225	CG32225	LAMP5	ENSG00000125869
FBgn0035538	DopEcR	GPR52	ENSG00000203737
FBgn0051660	smog	GPR158	ENSG00000151025
FBgn0033192	Corin	CORIN	ENSG00000145244
FBgn0266137	Dop1R2	ADRA1D	ENSG00000171873
FBgn0031697	CG14024	CHST10	ENSG00000115526
FBgn0035287	CG13937	CHST10	ENSG00000115526
FBgn0266137	Dop1R2	ADRA1B	ENSG00000170214
FBgn0266579	tau	MAPT	ENSG00000186868
FBgn0037408	NPFR	NPY5R	ENSG00000164129
FBgn0052311	zormin	CCDC141	ENSG00000163492
FBgn0037408	NPFR	NPY4R	ENSG00000204174
FBgn0037408	NPFR	NPY1R	ENSG00000164128
FBgn0000038	nAChRbeta1	CHRNA1	ENSG00000138435
FBgn0261090	Sytbeta	SYT6	ENSG00000134207
FBgn0261090	Sytbeta	SYT10	ENSG00000110975
FBgn0261090	Sytbeta	SYT9	ENSG00000170743
FBgn0028370	kek3	LRRC4C	ENSG00000148948
FBgn0029708	CG3556	CBLIF	ENSG00000134812
FBgn0029708	CG3556	TCN1	ENSG00000134827
FBgn0035477	CG14982	FAM110D	ENSG00000197245
FBgn0035477	CG14982	FAM110B	ENSG00000169122
FBgn0035477	CG14982	FAM110A	ENSG00000125898
FBgn0035477	CG14982	FAM110C	ENSG00000184731
FBgn0262018	CadN2	CDH1	ENSG00000039068
FBgn0262869	Gfrl	GFRAL	ENSG00000187871
FBgn0035331	MsR1	GPR139	ENSG00000180269
FBgn0013467	igl	SPA17	ENSG00000064199
FBgn0085380	CG34351	RGS7BP	ENSG00000186479

FBgn0038295	Gyc88E	GUCY1B2	ENSG00000123201
FBgn0010114	hig	CR2	ENSG00000117322
FBgn0010114	hig	SELP	ENSG00000174175
FBgn0039862	kek6	LRFN1	ENSG00000128011
FBgn0028888	CG4168	IGFALS	ENSG00000099769
FBgn0034286	dpr13	CD86	ENSG00000114013
FBgn0014073	Tie	TEK	ENSG00000120156
FBgn0039862	kek6	LRRC24	ENSG00000254402
FBgn0053988	Mid1	NALF2	ENSG00000130054
FBgn0053988	Mid1	NALF1	ENSG00000204442
FBgn0035719	tow	C1orf21	ENSG00000116667
FBgn0034788	CG13532	ICOSLG	ENSG00000160223
FBgn0034788	CG13532	LOC102723996	ENSG00000277117
FBgn0029974	dpr14	CADM2	ENSG00000175161
FBgn0029974	dpr14	CADM4	ENSG00000105767
FBgn0037993	dpr15	CADM4	ENSG00000105767
FBgn0037993	dpr15	CADM2	ENSG00000175161
FBgn0037908	dpr5	CADM2	ENSG00000175161
FBgn0037908	dpr5	CADM4	ENSG00000105767
FBgn0261871	dpr2	CADM2	ENSG00000175161
FBgn0261871	dpr2	CADM4	ENSG00000105767
FBgn0260995	dpr21	CADM2	ENSG00000175161
FBgn0260995	dpr21	CADM4	ENSG00000105767
FBgn0033250	CG14762	NYX	ENSG00000188937
FBgn0040351	CG11638	CABP4	ENSG00000175544
FBgn0083950	side-VI	CADM2	ENSG00000175161
FBgn0259213	side-II	CADM2	ENSG00000175161
FBgn0086604	side-VIII	CADM2	ENSG00000175161
FBgn0250908	beat-VII	CD80	ENSG00000121594
FBgn0035010	CG13579	TAAR1	ENSG00000146399

FBgn0037736	side-VII	CADM2	ENSG00000175161
FBgn0037736	side-VII	PSG8	ENSG00000124467
FBgn0260793	2mit	PODNL1	ENSG00000132000
FBgn0052057	dpr10	LRIT3	ENSG00000183423

Appendix B: Top 332 human random gene orthologs

FlyBaseID	Fly Symbol	Human Symbol	Ensembl ID
FBgn0262619	DNAlig1	LIG1	ENSG00000105486
FBgn0039407	CG14544	TRMT61A	ENSG00000166166
FBgn0030482	CG1673	BCAT1	ENSG00000060982
FBgn0038056	CG5961	FBXO9	ENSG00000112146
FBgn0027338	Kap-alpha3	KPNA4	ENSG00000186432
FBgn0250788	beta-Spec	SPTBN1	ENSG00000115306
FBgn0035064	TyrRS-m	YARS2	ENSG00000139131
FBgn0266724	Trs20	TRAPPC2	ENSG00000196459
FBgn0037901	Exd2	EXD2	ENSG00000081177
FBgn0058045	CG40045	UBE2G1	ENSG00000132388
FBgn0027082	ProRS-m	PARS2	ENSG00000162396
FBgn0034046	tun	NTAQ1	ENSG00000156795
FBgn0028693	Rpn12	PSMD8	ENSG00000099341
FBgn0027090	GlnRS	QARS1	ENSG00000172053
FBgn0002719	Men	ME1	ENSG00000065833
FBgn0033762	ZnT49B	SLC30A9	ENSG00000014824
FBgn0250837	dUTPase	DUT	ENSG00000128951
FBgn0035383	CPT2	CPT2	ENSG00000157184
FBgn0027087	HisRS	HARS1	ENSG00000170445
FBgn0000588	esc	EED	ENSG00000074266
FBgn0027779	VhaSFD	ATP6V1H	ENSG00000047249
FBgn0032487	Ski6	EXOSC4	ENSG00000178896
FBgn0028737	eEF1beta	EEF1B2	ENSG00000114942
FBgn0029648	CG3603	HSD17B8	ENSG00000204228
FBgn0039869	CG1890	TBCA	ENSG00000171530
FBgn0026430	Grip84	TUBGCP2	ENSG00000130640
FBgn0036733	U4-U6-60K	PRPF4	ENSG00000136875
FBgn0030973	CG7332	MINDY3	ENSG00000148481

FBgn0038771	CG4390	ESD	ENSG00000139684
FBgn0085377	CG34348	TMEM68	ENSG00000167904
FBgn0016698	Ptpa	PTPA	ENSG00000119383
FBgn0036921	RhoGDI	ARHGDIA	ENSG00000141522
FBgn0027580	PCB	PC	ENSG00000173599
FBgn0031600	CG3652	YIPF6	ENSG00000181704
FBgn0028691	Rpn9	PSMD13	ENSG00000185627
FBgn0036762	CG7430	DLD	ENSG00000091140
FBgn0037370	CG1236	GRHPR	ENSG00000137106
FBgn0046296	DEF8	DEF8	ENSG00000140995
FBgn0035839	CG7550	ADO	ENSG00000181915
FBgn0039357	CG4743	SLC25A26	ENSG00000144741
FBgn0053544	Vkor	VKORC1	ENSG00000167397
FBgn0000317	ck	MYO7A	ENSG00000137474
FBgn0017558	Pdk	PDK3	ENSG00000067992
FBgn0030230	Rph	RPH3A	ENSG00000089169
FBgn0261396	Rpn3	PSMD3	ENSG00000108344
FBgn0003415	skd	MED13	ENSG00000108510
FBgn0035916	GAPsec	TBC1D13	ENSG00000107021
FBgn0038662	Mpc1	MPC1	ENSG00000060762
FBgn0037087	CG7519	PBDC1	ENSG00000102390
FBgn0283450	Glo1	GLO1	ENSG00000124767
FBgn0027546	CG4766	MAB21L2	ENSG00000181541
FBgn0039636	Atg14	ATG14	ENSG00000126775
FBgn0022213	Cse1	CSE1L	ENSG00000124207
FBgn0086384	Mer	NF2	ENSG00000186575
FBgn0038013	CG10038	FAM172A	ENSG00000113391
FBgn0037021	CG11399	PCIF1	ENSG00000100982
FBgn0031639	mRpS2	MRPS2	ENSG00000122140
FBgn0037603	CG11753	SYS1	ENSG00000204070

FBgn0035206	sturkopf	LDAH	ENSG00000118961
FBgn0038115	CG7966	SELENBP1	ENSG00000143416
FBgn0259482	Mob3	MOB3A	ENSG00000172081
FBgn0260985	RfC4	RFC2	ENSG00000049541
FBgn0037466	CG1965	PAXBP1	ENSG00000159086
FBgn0005655	PCNA	PCNA	ENSG00000132646
FBgn0036741	anchor	GPR155	ENSG00000163328
FBgn0261276	Opa1	OPA1	ENSG00000198836
FBgn0037377	CG1218	HPF1	ENSG00000056050
FBgn0085224	CG34195	METTL6	ENSG00000206562
FBgn0033699	RpS11	RPS11	ENSG00000142534
FBgn0040271	Sulf1	SULF1	ENSG00000137573
FBgn0020909	Rtc1	RCL1	ENSG00000120158
FBgn0000183	BicD	BICD2	ENSG00000185963
FBgn0262117	IntS3	INTS3	ENSG00000143624
FBgn0029833	MCTS1	MCTS1	ENSG00000232119
FBgn0035965	Use1	USE1	ENSG00000053501
FBgn0030177	CG2972	NOB1	ENSG00000141101
FBgn0263606	Hsc20	HSCB	ENSG00000100209
FBgn0260962	pic	DDB1	ENSG00000167986
FBgn0034915	eIF6	EIF6	ENSG00000242372
FBgn0030645	Alg14	ALG14	ENSG00000172339
FBgn0260859	Bet3	TRAPPC3	ENSG00000054116
FBgn0023388	Dap160	ITSN1	ENSG00000205726
FBgn0264672	Eogt	EOGT	ENSG00000163378
FBgn0013954	Fkbp12	FKBP1B	ENSG00000119782
FBgn0035603	Pfdn4	PFDN4	ENSG00000101132
FBgn0024947	NTPase	ENTPD5	ENSG00000187097
FBgn0051156	CG31156	SRBD1	ENSG00000068784
FBgn0262467	Scox	SCO1	ENSG00000133028

FBgn0053198	pen-2	PSENEN	ENSG00000205155
FBgn0038597	CG8064	WDR3	ENSG00000065183
FBgn0038660	Sgsh	SGSH	ENSG00000181523
FBgn0034243	Ns2	GNL2	ENSG00000134697
FBgn0032720	mRpL13	MRPL13	ENSG00000172172
FBgn0003941	RpL40	UBA52	ENSG00000221983
FBgn0039713	RpS8	RPS8	ENSG00000142937
FBgn0020513	Paics	PAICS	ENSG00000128050
FBgn0260639	gammaTub23C	TUBG1	ENSG00000131462
FBgn0037647	RagA-B	RRAGA	ENSG00000155876
FBgn0028499	CG7985	HEXD	ENSG00000169660
FBgn0041147	ida	ANAPC5	ENSG00000089053
FBgn0264492	CkIIalpha	CSNK2A1	ENSG00000101266
FBgn0031566	CG2818	GPCPD1	ENSG00000125772
FBgn0035890	CG13667	NDOR1	ENSG00000188566
FBgn0039877	Mccc1	MCCC1	ENSG00000078070
FBgn0001986	Mtr4	MTREX	ENSG00000039123
FBgn0037647	RagA-B	RRAGB	ENSG00000083750
FBgn0038524	sll	SLC35B2	ENSG00000157593
FBgn0032061	CG9314	CAT	ENSG00000121691
FBgn0024887	kin17	KIN	ENSG00000151657
FBgn0042083	Mccc2	MCCC2	ENSG00000131844
FBgn0040273	Spt5	SUPT5H	ENSG00000196235
FBgn0032256	RluA-2	RPUSD2	ENSG00000166133
FBgn0038740	CG4562	ABCC4	ENSG00000125257
FBgn0035082	CG2811	GGACT	ENSG00000134864
FBgn0037756	CG8507	LRPAP1	ENSG00000163956
FBgn0033544	CG7220	UBE2W	ENSG00000104343
FBgn0032732	EMC5	MMGT1	ENSG00000169446
FBgn0039465	Tsp97E	TSPAN13	ENSG00000106537

FBgn0010215	alpha-Cat	CTNNA2	ENSG00000066032
FBgn0020388	Gcn5	KAT2A	ENSG00000108773
FBgn0024957	Irp-1B	ACO1	ENSG00000122729
FBgn0004924	Top1	TOP¬†1.00	ENSG00000198900
FBgn0003360	sesB	SLC25A4	ENSG00000151729
FBgn0027872	rdgBbeta	PITPNC1	ENSG00000154217
FBgn0039293	Alg9	ALG9	ENSG00000086848
FBgn0023169	AMPKalpha	PRKAA2	ENSG00000162409
FBgn0033055	Tbce	TBCE	ENSG00000284770
FBgn0034033	CG8204	ZNHIT3	ENSG00000273611
FBgn0034707	MED16	MED16	ENSG00000175221
FBgn0013997	Nrx-IV	CNTNAP2	ENSG00000174469
FBgn0034646	Rae1	RAE1	ENSG00000101146
FBgn0034259	P32	C1QBP	ENSG00000108561
FBgn0024994	Ugalt	SLC35A2	ENSG00000102100
FBgn0051866	Ada1-2	TADA1	ENSG00000152382
FBgn0034532	CG13436	CFAP206	ENSG00000272514
FBgn0032475	Sfmbt	MBTD1	ENSG00000011258
FBgn0033669	PI31	PSMF1	ENSG00000125818
FBgn0039427	CG5447	GOLGA7	ENSG00000147533
FBgn0004643	Zw10	ZW10	ENSG00000086827
FBgn0036614	Golgin104	CCDC186	ENSG00000165813
FBgn0262954	Rpb12	POLR2K	ENSG00000147669
FBgn0030311	CG11699	TMEM242	ENSG00000215712
FBgn0037881	GCC88	GCC1	ENSG00000179562
FBgn0010215	alpha-Cat	CTNNA1	ENSG00000044115
FBgn0023167	SmD3	SNRPD3	ENSG00000100028
FBgn0053057	CG33057	TRPT1	ENSG00000149743
FBgn0051717	CG31717	PLPP6	ENSG00000205808
FBgn0052590	CG32590	BORCS8	ENSG00000254901

FBgn0286788	Orc1	ORC1	ENSG00000085840
FBgn0029755	Sas10	UTP3	ENSG00000132467
FBgn0260407	mRpS23	MRPS23	ENSG00000181610
FBgn0034002	CG8079	AGGF1	ENSG00000164252
FBgn0036770	Prestin	SLC26A5	ENSG00000170615
FBgn0085436	Not1	CNOT1	ENSG00000125107
FBgn0037094	CG7611	WDR26	ENSG00000162923
FBgn0039427	CG5447	GOLGA7B	ENSG00000155265
FBgn0000114	bru1	CELF1	ENSG00000149187
FBgn0031977	baf	BANF1	ENSG00000175334
FBgn0039561	mfrn	SLC25A37	ENSG00000147454
FBgn0022361	Pur-alpha	PURA	ENSG00000185129
FBgn0034876	wmd	STRAP	ENSG00000023734
FBgn0032456	MRP	ABCC3	ENSG00000108846
FBgn0039152	Root	CROCC	ENSG00000058453
FBgn0038300	Mau2	MAU2	ENSG00000129933
FBgn0086613	Ino80	INO80	ENSG00000128908
FBgn0025621	CG16989	HEATR6	ENSG00000068097
FBgn0053506	CG33506	TMEM177	ENSG00000144120
FBgn0037109	MED1	MED1	ENSG00000125686
FBgn0031238	CG3645	DUS1L	ENSG00000169718
FBgn0283666	Rap2l	RAP2B	ENSG00000181467
FBgn0283531	Duox	DUOX2	ENSG00000140279
FBgn0033738	DUBAI	USP35	ENSG00000118369
FBgn0284251	l(2)05287	WDR75	ENSG00000115368
FBgn0035415	CG14966	C15orf40	ENSG00000169609
FBgn0261550	CG42668	OSBPL8	ENSG00000091039
FBgn0051721	Trim9	TRIM9	ENSG00000100505
FBgn0026619	Taz	TAFAZZIN	ENSG00000102125
FBgn0033738	DUBAI	USP38	ENSG00000170185

FBgn0026160	tna	ZMIZ2	ENSG00000122515
FBgn0283666	Rap2l	RAP2C	ENSG00000123728
FBgn0027841	CstF64	CSTF2T	ENSG00000177613
FBgn0027841	CstF64	CSTF2	ENSG00000101811
FBgn0039663	CG2321	C2orf42	ENSG00000115998
FBgn0038474	mRpS11	MRPS11	ENSG00000181991
FBgn0001222	Hsf	HSF1	ENSG00000185122
FBgn0038582	CG7988	SRRD	ENSG00000100104
FBgn0016696	Pitslre	CDK11B	ENSG00000248333
FBgn0026401	Nipped-B	NIPBL	ENSG00000164190
FBgn0032428	CG6405	CIBAR1	ENSG00000188343
FBgn0032407	Pex19	PEX19	ENSG00000162735
FBgn0042693	wrd	PPP2R5D	ENSG00000112640
FBgn0262124	uex	CNNM2	ENSG00000148842
FBgn0003390	shf	WIF1	ENSG00000156076
FBgn0036366	JMJD7	JMJD7	ENSG00000243789
FBgn0016696	Pitslre	CDK11A	ENSG00000008128
FBgn0260462	CG12163	CTSF	ENSG00000174080
FBgn0261550	CG42668	OSBPL5	ENSG00000021762
FBgn0032683	kon	CSPG4	ENSG00000173546
FBgn0035540	Syx17	STX17	ENSG00000136874
FBgn0031344	CG7420	SERGEF	ENSG00000129158
FBgn0031730	CG7236	CDKL1	ENSG00000100490
FBgn0043002	Chrac-14	POLE3	ENSG00000148229
FBgn0030400	Pits	IRF2BPL	ENSG00000119669
FBgn0031799	Pez	PTPN21	ENSG00000070778
FBgn0265174	PIG-V	PIGV	ENSG00000060642
FBgn0038927	CG6015	CDC40	ENSG00000168438
FBgn0025808	Rad17	RAD17	ENSG00000152942
FBgn0051092	LpR2	LDLR	ENSG00000130164

FBgn0051092	LpR2	VLDLR	ENSG00000147852
FBgn0036684	CG3764	FNIP1	ENSG00000217128
FBgn0036684	CG3764	FNIP2	ENSG00000052795
FBgn0083992	Mkp	DUSP19	ENSG00000162999
FBgn0020224	Cbl	CBL	ENSG00000110395
FBgn0031799	Pez	PTPN14	ENSG00000152104
FBgn0028372	isopeptidase-T-3	UBAC1	ENSG00000130560
FBgn0034694	Plekhm1	PLEKHM1	ENSG00000225190
FBgn0033350	CG8237	FAM8A1	ENSG00000137414
FBgn0025111	Ant2	SLC25A6	ENSG00000169100
FBgn0050105	CG30105	RNASEH2C	ENSG00000172922
FBgn0035268	CG8001	DCAF8L2	ENSG00000189186
FBgn0035268	CG8001	DCAF8L1	ENSG00000226372
FBgn0020762	Atet	ABCG4	ENSG00000172350
FBgn0036853	mRpL21	MRPL21	ENSG00000197345
FBgn0051145	CG31145	FAM20C	ENSG00000177706
FBgn0036893	PIG-F	PIGF	ENSG00000151665
FBgn0052446	Atox1	ATOX1	ENSG00000177556
FBgn0265726	Nna1	AGBL2	ENSG00000165923
FBgn0011640	lark	RBM4B	ENSG00000173914
FBgn0003891	tud	TDRD6	ENSG00000180113
FBgn0031484	CG3165	TREX2	ENSG00000183479
FBgn0031484	CG3165	TREX1	ENSG00000213689
FBgn0040752	Prosap	SHANK1	ENSG00000161681
FBgn0037773	CG5359	DYNLT2B	ENSG00000213123
FBgn0083953	CG34117	FMC1	ENSG00000164898
FBgn0040871	CG12479	MICOS10	ENSG00000173436
FBgn0284408	trol	HSPG2	ENSG00000142798
FBgn0003117	pnr	GATA4	ENSG00000136574
FBgn0035179	CG12038	DIPK1C	ENSG00000187773

FBgn0022959	yps	YBX2	ENSG00000006047
FBgn0037022	CG11396	TRAPPC12	ENSG00000171853
FBgn0032488	CG16812	C19orf47	ENSG00000160392
FBgn0262582	cic	CIC	ENSG00000079432
FBgn0034086	CG8441	PRKRIP1	ENSG00000128563
FBgn0031188	CG14613	FAM210A	ENSG00000177150
FBgn0052392	Rsph3	RSPH3	ENSG00000130363
FBgn0260224	CG42498	LAGE3	ENSG00000196976
FBgn0002567	Rab32	RAB32	ENSG00000118508
FBgn0017581	Lk6	MKNK1	ENSG00000079277
FBgn0027073	Ugt49B1	UGT2B11	ENSG00000213759
FBgn0027073	Ugt49B1	UGT1A6	ENSG00000167165
FBgn0027073	Ugt49B1	UGT2B10	ENSG00000109181
FBgn0051076	CG31076	LRMDA	ENSG00000148655
FBgn0015524	otp	OTP	ENSG00000171540
FBgn0011766	E2f1	E2F1	ENSG00000101412
FBgn0004369	Ptp99A	PTPRG	ENSG00000144724
FBgn0035140	BORCS6	BORCS6	ENSG00000196544
FBgn0011766	E2f1	E2F2	ENSG00000007968
FBgn0011766	E2f1	E2F3	ENSG00000112242
FBgn0004882	orb	CPEB1	ENSG00000214575
FBgn0035443	CG12010	SPATA5L1	ENSG00000171763
FBgn0010620	CG10939	SLC9A3R1	ENSG00000109062
FBgn0040907	mRpL33	MRPL33	ENSG00000243147
FBgn0031191	Cp110	CCP110	ENSG00000103540
FBgn0035249	CG17249	PRCC	ENSG00000143294
FBgn0083981	RunxA	RUNX3	ENSG00000020633
FBgn0265184	CG44249	RNPC3	ENSG00000185946
FBgn0023091	dimm	BHLHA15	ENSG00000180535
FBgn0030947	CG6696	MEP1B	ENSG00000141434

FBgn0042630	Sox21b	SOX21	ENSG00000125285
FBgn0035842	CG7504	SETX	ENSG00000107290
FBgn0037944	CG6923	RNF111	ENSG00000157450
FBgn0022981	rpk	ASIC5	ENSG00000256394
FBgn0029907	Atx-1	ATXN1	ENSG00000124788
FBgn0262527	nsl1	KANSL1	ENSG00000120071
FBgn0037739	CG12948	C14orf119	ENSG00000179933
FBgn0001217	Hsc70-2	HSPA1A	ENSG00000204389
FBgn0001217	Hsc70-2	HSPA1B	ENSG00000204388
FBgn0053208	Mical	MICAL3	ENSG00000243156
FBgn0031228	ND-15	NDUFS5	ENSG00000168653
FBgn0023489	Pph13	ALX3	ENSG00000156150
FBgn0037007	CG5059	BNIP3L	ENSG00000104765
FBgn0038043	CG17202	MYCBP	ENSG00000214114
FBgn0032877	CG2617	RNF26	ENSG00000173456
FBgn0259219	CG42319	PDLIM2	ENSG00000120913
FBgn0263980	Stacl	STAC3	ENSG00000185482
FBgn0033961	ND-B15	NDUFB4	ENSG00000065518
FBgn0030893	RhoGAP16F	ARHGAP27	ENSG00000159314
FBgn0051038	CG31038	FAM89A	ENSG00000182118
FBgn0010323	Gsc	GSC	ENSG00000133937
FBgn0024993	CG2662	SAMD13	ENSG00000203943
FBgn0030617	CG9095	SELE	ENSG00000007908
FBgn0010622	DCTN3-p24	DCTN3	ENSG00000137100
FBgn0016641	PTP-ER	PTPN7	ENSG00000143851
FBgn0052202	CG32202	TFPT	ENSG00000105619
FBgn0264002	MsR2	GPR139	ENSG00000180269
FBgn0030617	CG9095	SELL	ENSG00000188404
FBgn0034420	CG10737	C2CD2	ENSG00000157617
FBgn0038601	CG18600	POLR1E	ENSG00000137054

FBgn0035665	Jon65Aiii	CTRB2	ENSG00000168928
FBgn0020906	Jon25Bi	CTRB2	ENSG00000168928
FBgn0035665	Jon65Aiii	CTRB1	ENSG00000168925
FBgn0027339	jim	MZF1	ENSG00000099326
FBgn0261269	conv	IGFALS	ENSG00000099769
FBgn0267431	Myo81F	MYO10	ENSG00000145555
FBgn0033781	CG13319	PSMG4	ENSG00000180822
FBgn0031597	CG17612	ZNF662	ENSG00000182983
FBgn0050443	Opbp	ZNF142	ENSG00000115568
FBgn0030410	Aven	AVEN	ENSG00000169857
FBgn0031675	CG9121	ASB18	ENSG00000182177
FBgn0040823	dpr6	CADM2	ENSG00000175161
FBgn0040823	dpr6	CADM4	ENSG00000105767
FBgn0021742	C901	DLK2	ENSG00000171462
FBgn0021742	C901	DLK1	ENSG00000185559
FBgn0038638	CG7702	CHAD	ENSG00000136457
FBgn0037120	CG11247	ZNF543	ENSG00000178229
FBgn0037120	CG11247	ZNF660-ZNF197	ENSG00000186448
FBgn0037120	CG11247	ZNF345	ENSG00000251247
FBgn0037120	CG11247	ZNF197	ENSG00000186448
FBgn0037120	CG11247	ZKSCAN7	ENSG00000196345
FBgn0037120	CG11247	ZNF675	ENSG00000197372
FBgn0037120	CG11247	ZNF431	ENSG00000196705
FBgn0037120	CG11247	ZNF780A	ENSG00000197782
FBgn0037120	CG11247	ZNF132	ENSG00000131849
FBgn0037120	CG11247	ZNF30	ENSG00000168661
FBgn0037120	CG11247	PRDM14	ENSG00000147596
FBgn0037120	CG11247	ZNF711	ENSG00000147180
FBgn0037120	CG11247	ZNF184	ENSG00000096654
FBgn0037120	CG11247	ZNF224	ENSG00000267680

FBgn0037120	CG11247	ZNF107	ENSG00000196247
FBgn0037120	CG11247	ZSCAN2	ENSG00000176371
FBgn0037120	CG11247	ZNF732	ENSG00000186777
FBgn0037120	CG11247	ZNF43	ENSG00000198521