Characterization of Drosophila melanogaster glial cell types using single-cell RNA sequencing data

Camille Couture
Supervised by Dr. Donald van Meyel

Integrated Program in Neuroscience, McGill University, Montréal
August 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Table of Contents

List of Figures

List of Abbreviations

Abbreviation	Explanation
ALH	After larval hatching
BP	Biological process
CNS	Central nervous system
DIOPT	DRSC integrative ortholog prediction tool
GEM	Gel bead in emulsion
GO	Gene ontology
LGB	Longitudinal Glioblast
MF	Molecular function
PC	Principal component
PG	Perineurial Glia
PCA	Principal component analysis
PNS	Peripheral nervous system
RP	Ribosomal protein
scRNAseq	Single cell RNA sequencing
SLC	Solute carrier
SPG	Subperineurial Glia
SNC	Système nerveux central
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier
VNC	Ventral nerve cord
vst	Variance-stabilizing transformation

Abstract

Glial cells in the central nervous system (CNS) are essential for neural development, metabolism, ion homeostasis, neuron excitability and, ultimately, for the execution of correct behaviors. However, the molecular and cellular mechanisms underlying these effects are incompletely understood. As yet, we are lacking a comprehensive understanding of i) the molecular features that characterize distinct types of glial cells, ii) heterogeneity among subpopulations within those types, iii) the specific relevance of those molecular features for glial cell function, and iv) the conservation of those features among species. To help address these important issues, my thesis research sought to provide a detailed molecular description of glial subtypes in the Drosophila brain and ventral nerve cord (VNC) at the level of the transcriptome. To do this, I used both published and unpublished single-cell RNA sequencing datasets acquired with the 10x Chromium microfluidics system from the CNS of Drosophila larvae and adults. I took advantage of RStudio's graphic interface to perform cluster analysis, assigned these clusters to particular glial cell types, then analyzed their transcriptome profiles to improve understanding of the molecular specialization of glial cells.

With particular focus on astrocytes, ensheathing glia and cortex glia, I compared the molecular marker profiles of each cell type in larvae and adults and then, using Gene Ontology (GO) to classify the molecular functions and biological processes reflected in these markers, I confirmed previous knowledge and shed new light on the manifold functions of glial cells. In addition, I have found that some of the heterogeneity found within glial cell types can be attributed to subpopulations that exist at distinct stages of the Drosophila life cycle, and in different regions of the CNS. With the help of our collaborators, I explored conservation of glial

cells between Drosophila and mice, identifying orthologous genes to the molecular markers identified in Drosophila glial and exploring the extent to which these orthologs are expressed in glial cells of the mouse brain. Together, the results of my MSc thesis research improves understanding of molecular profiles and heterogeneity among glial cells in Drosophila and will fuel further research into specialized functions of glial cells in the CNS.

Résumé

Les cellules gliales du système nerveux central (SNC) sont essentielles pour le développement neuronal, le maintien de l'homéostasie ionique, l'excitabilité et le métabolisme des neurones, ainsi que le comportement. Cependant, les mécanismes moléculaires et cellulaires sous-jacents sont encore mal compris. Pour l'instant, nous comprenons encore mal i) les caractéristiques moléculaires des différents sous-types de cellules gliales, ii) l'hétérogénéité au sein de ces sous-types, iii) la pertinence de ces caractéristiques moléculaires pour la fonction des cellules gliales, et iv) la conservation de ces caractéristiques entre espèces. Pour ces raisons, mes recherches ont cherché à apporter une description moléculaire détaillée des sous-types gliaux dans le SNC de la Drosophile, qui se compose du cerveau et du cordon nerveux ventral (CNV), via une analyse transcriptomique. Pour cela, j'ai utilisé deux bases de données obtenues par séquençage d'ARN de cellule unique (scRNAseq) (système microfluidique 10X Chromium) d'échantillons de SNC provenant de larves et d'adultes de Drosophile. J'ai utilisé l'interface graphique RStudio pour effectuer une analyse de groupement (cluster) des cellules, assigné ces groupements à des sous-types de cellules gliales, et analysé leurs profils transcriptomiques afin de mieux comprendre leurs spécificités moléculaires.

Je me suis concentrée sur 3 sous-types gliaux : les astrocytes, les cellules gliales engainantes et les cellules gliales corticales. J'ai d'abord comparé les profils de marqueurs moléculaires de chaque sous-type chez les larves et les adultes. J'ai utilisé ces marqueurs pour réaliser une analyse d'enrichissement de termes GO (Gene Ontology) et obtenu, pour chaque sous-type, un profil spécifique de fonctions moléculaires et de processus biologiques. L'exploration de ces différentes fonctions m'a permis d'une part, de confirmer des fonctionnalités

déjà connues pour ces cellules, et d'autre part, de découvrir de nouveaux processus spécifiques pour chaque sous-type, apportant ainsi un éclairage nouveau sur les multiples fonctions de ces cellules. Qui plus est, j'ai pu mettre en évidence une hétérogénéité au sein des sous-types gliaux, et montré qu'une partie de cette hétérogénéité résulte soit de différences liées au stade de vie de la drosophile, soit de différences régionales au sein du SNC. Enfin, avec l'aide de collaborateurs, j'ai exploré la conservation des types de cellules gliales et des fonctions entre la drosophile et les souris. Cette analyse comparative m'a permis d'identifier et d'explorer l'expression, dans les cellules gliales du cerveau de souris, de gènes orthologues aux marqueurs moléculaires des cellules gliales de drosophile. Ensemble, les résultats de ma recherche de maîtrise améliorent la compréhension des profils moléculaires et de l'hétérogénéité parmi les cellules gliales de la drosophile, et alimenteront des recherches supplémentaires sur les fonctions spécialisées des cellules gliales dans le SNC.

Acknowledgments

First and foremost, I would like to extend my gratitude and thanks to my supervisor Dr. Don van Meyel for believing in me and for the continued support inside and outside of the lab. I have learned so much and have grown as a scientist in these last two years and I will always be thankful for the opportunity.

I would secondly like to thank all of my lab-mates who have taught and supported me throughout, and whom I consider good friends; a particular thank you to Dr. Emilie Peco with whom I worked very closely on this project. Additionally, I want to thank Dr. Claire Marquilly, Dr. Eunjoo Cho, Muhammad Azman Akhter and Deirdre Hatton.

I am also thankful to our collaborators Dr. Ben Cocanougher from the Dr. Marta Zlatic lab at Janelia for providing us with the larval data set and for extracting glial cells from both larval and adult data sets. Additionally, Dr. Todd Farmer from the Dr. Keith Murai lab for help regarding the mammalian analyses.

I am grateful to my advisory committee members, Dr. Keith Murai and Dr. Brian Chen for their invaluable advice, guidance and encouragement, and to my IPN mentor Dr. Jesper Sjöström.

A further thanks is extended to the rest of the communities that I was a part of throughout my degree, such as the research community at the Centre for Research in Neuroscience (CRN), the Graduate Association of Neuroscience at McGill University (GSAN), the Ultimate frisbee team (MUT) and Women in Bio (WiB). It was a pleasure to be involved in such diverse environments that allowed me to grow inside and outside of academia.

Lastly but not least, I would like to thank all of my friends and family who have been understanding and unconditionally supportive throughout these last two years – particularly the stressful writing process during the Covid-19 pandemic. En particulier, je tiens à remercier mes sœurs, Meghan et Alissa Couture ainsi que mes parents, Jean-François Couture et Caroline Lafond pour avoir été des sources constantes de motivation et d'encouragement.

Introduction

1. Known glial cell functions in the central nervous system

The nervous system is the most complex biological system, controlling our physiology and behavior. Contributing to this complexity are multiple classes of neuronal and non-neuronal cells, which must cooperate to ensure proper nervous system development and function. To this day, the scientific research community has largely focused on elucidating the role of neurons and neural circuits in nervous system function, with less emphasis on the importance of non-neuronal cells such as glia.

Glial cells, or glia, are non-neuronal cells of the nervous system derived from neural precursor cells (Kessaris, Pringle, and Richardson 2008). Since their discovery over a century ago, thoughts on glial cells have shifted - they are now appreciated as essential components of the central and peripheral nervous systems. Glia were originally thought to function only in passive support of neurons – their name literally translating to "glue" - even though they account for more than half of the cells in the mammalian nervous system (Barres 2008). What's more, glia are known to be essential for neural development, central nervous system metabolism, ion homeostasis, neuronal excitability and behavior, spanning multiple scales - from single synapses to whole circuits (Brown and Ransom 2007; Clasadonte et al. 2013; Fields 2006; Halassa et al. 2009; Haydon et al. 2009; Haydon and Carmignoto 2006; Panatier et al. 2006). In addition, glial cells can release neurotransmitters such as glutamate to regulate neuronal activity in the brain and their dysfunction has been associated to many neurological disorders making them potential targets for novel treatments (Angulo et al. 2004). As we learn more about the typical functions of glia, their

importance in everyday functioning and also diseases of the nervous system, is becoming increasingly apparent.

In mammals, glia can be categorized into four major subtypes, three of which are found in the central nervous system (CNS): astrocytes, oligodendrocytes and microglia, and one in the peripheral nervous system (PNS): Schwann cells. Astrocytes are the most abundant cell type in the vertebrate CNS. They have diverse roles which span CNS development and plasticity (Boulanger and Shatz 2004), control of synaptic transmission (Shigetomi et al. 2008; Pascual et al. 2005), regulation of blood flow, energy and metabolism (Sofroniew and Vinters 2010), formation of the blood-brain barrier (Abbott, Ronnback, and Hansson 2006) and regulation of circadian rhythms (Jackson 2011). Astrocytes have been implicated in neurodegenerative diseases where their activation or dysfunction contribute to neurological disorders such as epilepsy, stroke and Alzheimer's disease (Acosta, Anderson, and Anderson 2017). Astrocytes are characterized, in part, by the fine processes they extend near both the pre-synaptic terminals of axons and the post-synaptic specializations of dendrites (Allen and Eroglu 2017). This proximity to neuronal synapses of neurons has made them an exciting avenue for studies of glial biology because it gives them direct opportunity to modulate neural activity.

Another important glial cell is the oligodendrocyte, which forms flat, sheath-like processes around axons that differentiate into an insulating sheath called myelin, enabling saltatory conduction of nerve impulses for transmission of information to be quick and efficient (Pereanu, Shy, and Hartenstein 2005). Their precursors (polydendrocytes) express the chondroitin sulfate proteoglycan: NG2, which is found expressed in 5-10% of all glia in the developing and adult CNS

(Trotter, Karram, and Nishiyama 2010). These cells become activated in response to a wide variety of insults and could play a primary role in pathogenesis by contributing to myelin repair (Nishiyama, Suzuki, and Zhu 2014). NG2 cells are particularly impressive because they expand their population by extensive self-renewal, even after their peak proliferation, and they retain their proliferative ability throughout life (Nishiyama, Suzuki, and Zhu 2014; Trotter, Karram, and Nishiyama 2010). In the PNS, Schwann cells take the place of oligodendrocytes as the myelinating glia (Aguayo et al. 1976).

Finally, microglia, the myeloid-derived resident macrophages of the brain, are involved in immune surveillance and responses to environmental stress within the CNS (Salter and Stevens 2017). In addition to immune receptors, microglia express multiple receptors for neurotransmitters released by neurons, allowing them to monitor and respond to neuronal activity by influencing synaptic plasticity and sculpting dendritic spine density (Pocock and Kettenmann 2007).

Together, distinct glial cell sub-types in mammals contribute to the proper functioning of the nervous system. Glia exhibit remarkably similar morphologies and functions in *Mus musculus* (Xu et al.), *Drosophila melanogaster* (fruit fly), *Danio rerio* (zebrafish), *Caenorhabditis elegans* (nematode) and other phyla such as birds and amphibians - suggesting that important glial functions are likely to be highly conserved (Xie et al. 2019; Muthukumar, Stork, and Freeman 2014; Colon-Ramos, Margeta, and Shen 2007).

2. Drosophila as a genetic model organism to study glial cells

Drosophila is particularly well suited for studying glia biology, with a genome one-tenth the size of mammals, Drosophila now come with a wealth of genetic tools needed to dissect complex cellular functions (Venken 2011; Schneider-Mizell et al. 2016; Ohyama et al. 2015; Li et al. 2014; Diao et al. 2015). Fruit flies are fecund, they have a short life cycle and are easily cultured and maintained in a laboratory. Most importantly for my project, the Drosophila nervous system contains only a limited number of glial cells (10%) while maintaining diversity and being involved in sophisticated behaviors, which can each be readily identified based on sets of molecular markers and on their position within the CNS (Yildirim et al. 2019).

The life cycle of Drosophila lasts approximately ten days under optimal conditions. Beginning as an embryo, stereotyped sets of neuroblasts generate small lineages of primary neurons and glial cells that differentiate to form the functional larval CNS. Larvae have a CNS that is composed of two brain lobes and a ventral nerve cord (VNC) (Figure 1A). It is organized into an outermost cortex layer where neuronal and glial cell bodies are located, and an innermost layer of neuropil where synapses are formed (Figure 1A). Larvae undergo three instars, or molts - occurring approximately 24h (1st instar), 48h (2nd instar) and 72 hours (3rd instar) after the larvae have hatched. In the larval CNS, some neuroblasts resume proliferating and produce large lineages of secondary neurons that invade the neuropil but do not differentiate any further (Dumstrei, Wang, and Hartenstein 2003). At the onset of metamorphosis, 3rd instar larvae undergo pupariation, during which secondary neurons form dendritic and axonal branches that become integrated into circuits within the developing adult CNS that also involve remodeled primary

neurons (Pereanu, Shy, and Hartenstein 2005). In this light, Drosophila is a fitting subject to study the neuroanatomy and development of the CNS, allowing researchers to benefit from its advanced and cost-effective genetic toolkit and relatively simple cellular composition and organization, which supports a range of innate and learned behaviors.

3. Glial cell types in the CNS of Drosophila melanogaster

In the Drosophila CNS, glial cells largely fall into four classes, three of which will be our focus for this thesis (highlighted in Figure 1B). Firstly, neuropil-associated glia, which include astrocytes (also referred to as astrocyte-like glia) and ensheathing glia. Both are located at the interface between the cortex and neuropil, forming a sheath around the neuropil compartments, as well as major tracts of neurites (Hoyle, Williams, and Phillips 1986). Second, cortex glia (also called cell-body glia) encapsulate neuronal cell bodies in the cortex region. Third are the surface glia, which extend sheath-like processes to wrap around the entire brain and are interconnected by septate junctions forming a 'blood-brain-barrier'. Surface glia can be further divided into perineurial and subperineurial glia. In the peripheral nervous system, sensory and motor nerves that exit the nerve cord are enveloped by perineurial glia and subperineurial glia and are infiltrated by wrapping glia that ensheathe axons (Stork et al. 2008). Each of these broad classes will be discussed in more detail below. Interestingly, however not a focus of my thesis, are the midline glia, an uncommon subtype derived from mesectoderm that enwrap axons at the CNS midline.

Larval glial cells are generated during embryogenesis and their numbers increase slowly during the first half of larval development. Their growth becomes more rapid during the third instar due to mitosis of differentiated glia and the proliferation of neuroblasts – where the increase

of cell number during larval development is shown to be strongest in cortex glia (15-fold), followed by neuropil glia (5-fold) and surface glia (3-fold) (Pereanu, Shy, and Hartenstein 2005). Some larval glial cells persist in the adult CNS while others are generated anew by gliogenesis occurring during the pupal stage (Awasaki et al. 2008; Enriquez et al. 2018; Kato, Orihara-Ono, and Awasaki 2020). Specifically, some larval neuropil glia die during metamorphosis and are replaced from larval neuro-glioblasts to form adult neuropil-associated glia, whereas surface and cortex glia remain – proliferating to form their adult populations (Omoto, Lovick, and Hartenstein 2016).

3.1. Neuropil-associated glial cells: astrocytes and ensheathing glia.

At the heart of the Drosophila nervous system lies the neuropil, which contains dendrites, axons, and synapses. Glial cells that come into contact with the neuropil are called neuropil-associated and include astrocytes and ensheathing glia.

During development, neuropil-associated glial cells originate from Longitudinal Glioblast (LGB) precursor cells in the embryonic neuroectoderm (Beckervordersandforth et al. 2008). Expression of the transcription factor glial cells missing (gmc) in LGB precursors promotes the glial cell fate (Hosoya et al. 1995; Jones et al. 1995; Vincent, Vonesch, and Giangrande 1996). This transcription factor regulates reversed polarity (repo) gene expression - a DNA-binding homeoprotein used as a pan-glial marker at all stages of Drosophila development (Halter et al. 1995; Xiong et al. 1994). In each segment of the larval VNC, the neuropil-associated glia arise from a common lineage, where nine embryonic LGBs differentiate into three ensheathing glia and six astrocytes (Peco et al. 2016). Interestingly, each astrocyte is allocated to cover consistent regions of the neuropil, a feature similar to the regional allocation of astrocytes that has been observed in

the spinal cord of rodents (Bushong, Martone, and Ellisman 2004; Distler, Dreher, and Stone 1991; Ogata and Kosaka 2002)

Astrocytes are quite similar between Drosophila and mammals in their morphology, molecular composition and functionality within the CNS. Firstly, Drosophila larval astrocytes extend processes into the neuropil, infiltrating non-overlapping territories, similar to the "tiling" of astrocytes in mammals (Peco et al. 2016). Here they can be electrically coupled with one another (MacNamee et al. 2016; Huang, Ng, and Jackson 2015; Ng et al. 2016) and exhibit endogenous spontaneous Ca^{2+} oscillatory activity established by fast, recurrent microdomain Ca^{2+} fluctuations (Zhang, Ormerod, and Littleton 2017). Astrocytes participate in neurotransmitter homeostasis and express a number of transporters. Experiments from our lab have shown high expression of Excitatory Amino Acid Transporter 1 (Eaat1, corresponding to mouse GLAST) in astrocytes, and that loss-of-function of Eaat1 causes distinct locomotor deficits at larval stages (Stacey 2010) – demonstrating the importance of astrocytes in neural function. Other important factors found in astrocytes include: GABA transporter (Gat), the transcription factor Prospero (Lefevre et al. 2000) and glutamine synthetase 2 (Gs2) (Freeman et al. 2003; Stork et al. 2014; Peco et al. 2016; Stacey et al. 2007). Gs2 enables Eaat1-transported glutamate to be converted into glutamine, allowing it to be safely recycled back into neurons. This glutamate recycling function of astrocytes is conserved between mammals and Drosophila (Yang et al. 2009; Benediktsson et al. 2012; Devaraju et al. 2013).

Contrary to astrocytes, ensheathing glia, have flattened processes that do not infiltrate the neuropil but rather enclose it – subdividing the neuropil compartments within brain lobes (Hartenstein 2011) (Figure 1B). Ensheathing glia also play critical roles in proper CNS

morphogenesis, where their ablation impairs the formation of axon tracts and alters axonal trajectories (Dumstrei, Wang, and Hartenstein 2003; Pereanu et al. 2010; Spindler et al. 2009). In the adult brain, ensheathing glia potently respond to injury-induced axon degeneration and are capable of clearing debris of damaged axons from the neuropil (MacDonald et al. 2006). Altogether, this data illustrates the importance of neuropil-associated glia in the regulation and proper functioning of the Drosophila CNS.

3.2. Cortex glia

Cortex glia are one of the least studied glial sub-type in Drosophila. As their name implies, they reside in the neuronal cell cortex where they encapsulate individual cell bodies of neurons – forming a scaffold and providing a link between the subperineurial glia and neurons (Pereanu, Shy, and Hartenstein 2005) (Figure 1B). Each cortex glial cell can envelop around 100 neuronal cell bodies, forming a "trophospongium" – a honeycomb-like structure of supportive glial membranes (Awasaki et al. 2008). Cortex glia are highly branched and undergo extensive rearrangements during development where their processes play essential roles in stabilizing neuronal cell body positions during larval development (Pereanu, Shy, and Hartenstein 2005). Cortex glia also exhibit spontaneous and activity-dependent fluctuations in intracellular calcium levels, which is dependent on the glial-specific potassium dependent sodium/calcium exchanger Zydeco (zyd) (Melom and Littleton 2013). Due to their contact with the SPG as well as the Drosophila vasculature, it has been suggested that cortex glia might play roles as conduits for nutrient transport from the hemolymph as well as gas exchange (Pereanu et al. 2007).

3.3. Surface Glia

The nervous system (both central and peripheral) has to be metabolically separated from the rest of the body, classically done by a blood–brain barrier (Abbott, Ronnback, and Hansson 2006; Carlson et al. 2000; Tietz and Engelhardt 2015). The metabolic barrier in Drosophila, serves to separate the nervous system from the hemolymph (insect equivalent of blood) and is made up of two types of surface glia (Freeman and Doherty 2006). The outer-most layer of the barrier is composed of the perineurial glial (PG) cells – small, voluminous cells with many protrusions but that lack contact with neurons (Awasaki et al. 2008; Stork et al. 2008). It has been previously speculated that cells considered to be PG might be hemocytes – motile cells acting like macrophages (Tepass and Hartenstein 1994). Directly below the PG lies the subperineurial glia (SPG) - very large flat cells held tightly by septate junctions (Bainton et al. 2005; Carlson et al. 2000; Schwabe et al. 2005; Stork et al. 2008). Together PG and SPG establish a strong diffusion barrier. Developmentally, surface glia also appear to produce signals required for neuroblast proliferation and contain many ATP-binding cassettes, solute carrier transporters, cell adhesion molecules, metabolic enzymes and other signaling molecules that allow them to interreact with their environment (Ebens et al. 1993; DeSalvo et al. 2014).

In terms of abundance, astrocytes, ensheathing glia, cortex glia, and surface glia make up 34%, 27%, 20% and 19% of all glia in the adult CNS respectively (Kremer et al. 2017). My research project will focus primarily on astrocytes, ensheathing glia, and cortex glia.

4. Single Cell RNA Sequencing to characterize glial cell types.

Understanding the nervous system undoubtedly involves identifying, characterizing and comprehending its cellular elements, and one way to do this is by profiling gene expression of single cells within their appropriate cellular context. For neurons in the CNS, gene expression determines how they behave, how they are structured, and their connectivity (Croset, Treiber, and Waddell 2018). More than ever before, we can now correlate genes expression with information about cellular connectivity and activity thanks to recent advances in connectomics (Eichler et al. 2017; Helmstaedter et al. 2013; Jarrell et al. 2012; Ohyama et al. 2015; Takemura et al. 2013) and live imaging techniques (Ahrens et al. 2013; Grimm et al. 2017; Hildebrand et al. 2017; Lemon et al. 2015; Prevedel et al. 2014). Likewise, elucidation of gene expression profiles of glia in the CNS is an important way to characterize and increase our understanding of glial cell function and heterogeneity.

Traditionally, identifying cell types has relied on meticulous histochemical classification. Patterns of differentially expressed genes have often been used to discriminate cells from one-another, however only a few genes are known to be cell-type specific, making current cell identification imprecise and potentially inconsistent. However, high-throughput single-cell RNA sequencing (scRNA-seq) now offers an alternative way forward, by providing a molecular-level identity for each cell via its RNA expression profile made up of thousands of genes (Figure 1C).

For my thesis research, we took advantage of data gathered by a method of scRNA-sequencing that uses microfluidic droplet-based technology (10x Chromium) to isolate single cells, and an oligonucleotide-based barcoding approach to tag RNA molecules from the same cell (Zheng

et al. 2017) (Figure1C). Other techniques provide average expression signal for an ensemble of cells based on the assumption that a tissue is homogenous, but scRNAseq allows us to distinguish the unique transcriptomes from distinct cell types in heterogeneous tissues. This technique is particularly important for my project to bring to light underlying differences in glial populations that might otherwise be missed from analyses of pooled cells.

Using this method, neural issues are dissected, dissociated and processed through micro-fluidics droplet-based technology where barcoded gel beds are encapsulated in oil droplets with individual cells called GEMS (gel bead in emulsion). This creates an isolated environment where RNA molecules from a single cell can be specifically tagged with the barcode from the beads during reverse transcription. cDNA libraries are then amplified via polymerase chain reaction (PCR) prior to sequencing. This approach allows us to know the number of unique RNA molecules per gene and the gene expression profiles from individual cells - giving us the resulting gene-by-cell expression matrices to delineate cell-to-cell diversity within populations at a high resolution.

This technique has already revealed striking heterogeneity in cell populations that was lost in other sequencing techniques such as bulk RNA sequencing. Examples where such work has contributed to our knowledge of Drosophila cell types are the adult brain atlas (Davie et al. 2018), the adult optic lobe atlas (Konstantinides et al. 2018), the adult midbrain atlas (Croset, Treiber, and Waddell 2018), the adult ventral nerve cord atlas (Allen et al. 2020), and the larval CNS atlas (Cocanougher et al. 2020). These labs have focused their efforts to deeply characterize distinct neuron populations, however, the purpose of my thesis research was to take advantage of this approach in order to characterize glial cells. In the future, this technique is likely to continue to

provide valuable knowledge about how gene products contribute to cell-specific functions and, ultimately, organismal behavior (Stahlberg et al. 2011; Rusnakova et al. 2013).

Goals of the project

With scRNAseq technology, we now have the ability to fully distinguish glial cells from one another at a molecular level providing us with the ability to characterize them with higher precision. The goal of my thesis research is to use scRNAseq to define and characterize glial cells in the Drosophila CNS at the level of the transcriptome using published and unpublished scRNAseq datasets at larval and adult stages. My project had four specific aims:

1. Distinguish glial cells via cluster analysis on scRNA-seq data from neural tissues at larval and adult stages.
2. Explore the molecular relationship between glial cell types from distinct stages of life and different regions of the CNS to define stage-specific specializations and investigate regional heterogeneity.
3. Consider the functional relevance of our findings via gene ontology analysis.
4. Examine the conservation of molecular features between glial cells in Drosophila and mice.

Methods

1. scRNAseq Data Sets used for cluster analysis

We took advantage of scRNAseq data sets from two sources. The larval data set was generated by the lab of Dr. Marta Zlatic using single cell RNA sequencing droplet-based microfluidics (10X Chromium) (Figure 1C). It consists of nearly 100,000 cells (100K) captured from the CNS of larvae (Cocanougher et al. 2020). This data set was collected at four distinct time points during larval development: 1 hour, 24h, 48h and 96h after larval hatching (ALH). It also involved dissections that provided three distinct CNS preparations: whole CNS, brain only, and VNC only. Our collaborators from the Zlatic lab were then able to identify glial cell clusters, extract them from the data set and send them to us for further processing.

The adult data set was generated in the lab of Dr. Stein Aerts using 10X Chromium technology and is comprised of 57,000 cells (57K) from 40 adult brains (20 females, 20 males), from 2 lab strains (DGRP-551 and w^{1118}) and at 8 time points (Davie et al. 2018). With a sequencing depth of 53,553 reads per cell and a sequence saturation rate of 81.5%, there were approximately 3,600 (6.4%) glial cells detected. This data set is currently publicly available on the web-based application Scope (http://scope.aertslab.org). Bioinformatic analysis of both the larval and adult data sets was performed using Cell Ranger software (Versions 1.3.1 and 2.0.0, 10x Genomics, Pleasanton, CA, USA), the Seurat R package (Satija et al. 2015) and custom software in R and Python. Our collaborator Ben Cocanaugher in the Zlatic lab extracted a subset of cells from the data sets enriched for glia and sent the data matrices to us for further processing.

2. scRNA-seq- pre-processing of data

Gene barcode matrices were analyzed in RStudio version 1.1.463 using Seurat package version 3.1.1 (Satija et al. 2015). For quality control, matrices included only cells with more than 200 genes and genes expressed in more than 3 cells. To remove potential instances where two cells might have been captured in the same GEM and thereby received an identical barcode, I next removed cells that had more than 2500 genes and greater than 5% mitochondrial gene counts. In all samples, UMI counts (or RNA counts) were log normalized with a scale factor of 10,000. Next I identified highly variable features by identifying genes with high cell-to-cell variation in the data set - choosing 2,000 genes that are highly variable for downstream analysis of principal components (PCs) using variance-stabilizing transformation ('vst'). The resulting outputs were scaled, and Unique Molecular Identifier (UMI) counts and mitochondrial gene contamination removed via regression using the ScaleData function. Principal component analysis (PCA) was performed with RunPCA and significant PCs were determined based on Jack Straw procedure and Elbow Plots.

3. Clustering

Clustering was determined by FindNeighbors using 10 dimensions followed by FindClusters using 10 dimensions, the original Louvain algorithm and a resolution parameter adjusted based on the biological relevance of the data (between 0.1-0.6). Clusters were visualized by Uniform Manifold Approximation and Projection (UMAP) with Seurat's RunUMAP function using 10 dimensions, 30 neighbors and a PCA reduction.

4. Cluster Marker Identification

Differentially expressed genes (markers) were identified using a number of criteria. First, using the FindAllMarkers function and a Wilcoxon Rank Sum test. Only genes that were significantly up-regulated were considered (Bonferroni adjusted p-value ≤ 0.01). Second, markers were selected to be at least 1.28-fold (log fold change = 0.25) overexpressed in the subtype of interest (when compared to other clusters), empirically chosen to balance the number of markers to identify and characterize each cluster with background levels of variation. Lastly, min.pct, a parameter to test genes that are detected in only a minimum fraction of cells in either of the two populations, was set to 0.1 and I returned only positive markers.

5. Comparisons of glial subtypes in Drosophila and mice:

The top 200 gene markers for our glial cell clusters of interest in Drosophila larvae and adults were isolated. Then, murine (*Mus musculus*) orthologs for each gene was identified using the Integrative Ortholog Prediction Tool (DIOPT) from DRSC Functional Genomics Resources (Hu et al. 2011). In DIOPT, all prediction tools were selected to find orthologs – based on phylogeny (TreeFam, Phylome, Ensembl Compara), sequence similarity (InParanoid, orthoMCL and OMA) and functional similarity (e.g. NetworkBLAST, IsoBase). DIOPT was set to return the "best match" orthologs when there was more than one match per input gene. These mouse genes were then used to query scRNAseq data published by the McCarroll lab (Saunders et al. 2018). Our collaborator Dr. Todd Farmer used an algorithm in R version 3.6.0 to determine which orthologs were differentially expressed among distinct neural cell types in the mouse brain. Markers were found using the FindAllMarkers function to isolate all the positive markers with a min.pct = 0.5 and logfc.threshold= 0.5. These marker genes were scaled and plotted on a heatmap of 500

representative mouse brain cells, then exported to identify orthologs differentially expressed in distinct classes of mouse glia.

6. Gene Ontology Analysis

Cytoscape is an open source software platform for visualizing complex networks of genes (Shannon et al. 2003). We used its ClueGO plug-in to decipher biological and molecular networks associated with Drosophila glial cell types identified with scRNAseq data. Our input genes were up regulated genes are defined using default Seurat marker parameters (logfc.threshold = 0.25, test.use = wilcox, min.pct = 0.1, min.diff.pct = -Inf). Enrichment is expressed as the logarithmic 10 of the p-value for the associated group of GO terms corrected with Benjamini-Hochberg. Only pV<0.05 were considered to be enriched significantly and were compared to a background of all Drosophila melanogaster genes. Statistical options selected included a pV correction with the Benjamin-Hochberg method to calculate a q-value (FDR) with a cut off of pV<0.01 (for Biological Processes) and 0.05 (for Molecular Functions). GO term options included an empirically determined minimum number of genes that mapped to a term, and the proportion (%) of genes with a shared term. For Molecular Function this was a minimum of 3 genes and a shared proportion of 4%. For Biological Process, this was a minimum of 4 genes and a shared proportion of 5%. We set GO term fusion to ON and selected all terms and pathways.

Results

1. Data processing with R to obtain glia-enriched data sets

Cluster analysis (or clustering) is a main task of exploratory mining of large data sets, and for my thesis research I used this approach to cluster cells into populations based on the relative similarity of their gene expression profiles. Using R to analyze scRNAseq data obtained via 10X (Chromium) sequencing, I chose an unsupervised clustering approach to remove bias and allow grouping based solely on similarities found through original Louvain algorithms. Following multiple iterations of different combinations of parameters, I found that setting resolution between 0.1-0.6 returned consistently satisfactory results for our single-cell datasets, depending on their size.

The starting larval CNS data set consisted of nearly 100,000 cells (100K, from Zlatic lab) and 13,000 genes (Figure 2A). Specifically, the larval 100k data was processed using the first 10 principal components and a resolution of 0.6, which resulted in 23 clusters with 8,531 nodes, 279,293 edges and a maximum modularity of 0.92. The starting adult brain data set consisted of 57,000 cells (57K, from Aerts lab) and 17,400 genes (Figure 3A). It was processed using the first 10 principal components and a resolution of 0.2, resulting in 12 clusters with 7,642 nodes, 251,832 edges and a maximum modularity of 0.96.

To compare data sets acquired from different labs (Zlatic and Aerts) we aimed to process them consistently. Since these CNS data sets contain many distinct cell types such as neurons, glia, immune cells, and trachea cells among others, I first sought to enrich for glia from this large heterogeneous pool. Both data sets went through an initial enrichment for glial cells by our collaborator Dr. Ben Cocanaugher in the Zlatic lab. His enrichment protocol left the larval data set with a matrix of 8,566 cells and 12,068 genes (Figure 2B) and the adult data set with 7,910 cells

and 12,925 genes (Figure 3B). Following my quality control and pre-processing steps, the larval data set contained 9,625 genes in 8,531 cells (8.5K data set), which grouped into 23 clusters (Figure 2C). After the same processing, the adult data set contained 10,839 genes expressed in 7,642 cells (7.6K data set) and grouped into 12 clusters (Figure 3C).

To avoid removing glial cells in the initial extraction, Dr. Cocanaugher was conservative with his enrichment criteria, however, at this stage we wanted to further enrich these data sets for glial cells by removing any neuronal cell populations that might remain. To do this, I used Seurat to identify genes that were differentially expressed between cells in different clusters. I then surveyed these gene markers of clusters for robust expression of neuron-specific markers known from the literature: neuronal Synaptobrevin (nSyb: an adaptive protein for vesicular release) (DiAntonio et al. 1993), embryonic lethal abnormal vision (elav: a neuron-selective RNA-binding protein) (Campos et al. 1987; Robinow and White 1988), paralytic (para: an alpha subunit of a voltage-gated sodium channel) (Piggott et al. 2019) and bruchpilot (brp: a constituent protein of pre-synaptic active zones) (Wagh et al. 2006). It was not enough for a cluster to show a lack of neuronal markers, it evidently also had to show an expression of glial markers such as the long non-coding RNA MRE16 (lncRNA:CR34335) (Davie et al. 2018), reversed polarity (repo) - a DNA-binding homeoprotein used as a pan-glial marker at all stages of Drosophila development (Halter et al. 1995; Xiong et al. 1994) and the Na^+/K^+ transporting ATPase nervana 2 (nrv2) (Sun and Salvaterra 1995).

In the larval 8.5k data set, three clusters (numbered 3, 6 and 11) had a combination of higher expression of neuronal markers brp, elav, nSyb and para, and low expression of glial markers (MRE16, repo and nrv2) (Figure 2D). The retention of neurons in this data confirmed that

Dr. Cocanaugher's initial glial enrichment from the complete 100k data was rather conservative and unlikely to have inadvertently cut glial populations from the data. These neuronal populations were then removed from the data, and the remaining 20 non-neuronal populations were re-clustered using 15 principal components, a resolution of 0.2 and Louvain algorithm to achieve an enriched larval glia data set that resulted in 13 clusters with 7,094 nodes, 247,274 edges and a maximum modularity of 0.96 (Figure 2E). Comprised of 7,094 cells, this glia-enriched data set is hereafter referred to as **larval 7K**. When screened for neuronal and glial markers, the larval 7K data set no longer showed high expression of neuronal markers, but rather had higher expression of glial markers, confirming the enrichment (Figure 2F).

Using the same approach for the adult data set, we identified and removed cells in four neuronal clusters (numbered 0, 5, 6 and 10) from the adult 7.6K data set because they expressed higher levels of neuronal markers brp, elav, nSyb and para, and had low expression of glial markers MRE16, repo and nrv2 (Figure 3D). Re-clustering of the remaining cells using the top 20 principal components and a resolution of 0.1 resulted in 10 clusters with 4,411 nodes, 139,458 edges and a maximum modularity of 0.97. Comprised of 4,411 high quality cells, this glia-enriched data set is hereafter called **adult 4.4K**. (Figure 3E). In the adult 7.6K data set, we noticed that cluster 3 had high expression of both neuronal and glia markers. In addition, cluster 11 had low expression of all glial markers and neuronal markers other than nSyb (Figure 3D). We did not remove these clusters because they did not meet our criteria for differentiated neurons (a combination of high neuronal and low glial gene markers), which is also why the 4.4K data set has two clusters with these patterns (clusters 2 and 8 respectively) (Figure 3F). The 4.4K data set showed high expression of glial markers, confirming the enrichment (Figure 3F).

In summary, using Seurat to re-cluster data from which non-glial cell types had been removed first by a glial-enrichment protocol and then by selective removal of residual cells expressing neuronal markers, the larval 7K data gave 13 clusters and adult 4.4K data gave 10 clusters. These clusters were again screened for the same 4 neuronal and 3 glial markers noted above, and we found that every cluster in the larval 7K data set had low expression of neuronal markers, and that all but clusters 8-12 had high expression of glial markers (Figure 2F). We believe clusters 8-12 are glia-like because they had low expression of neuronal markers, and higher expression of at least one glial marker – and so I kept them for further downstream analyses. For the adult 4.4K data, clusters largely had high expression of glial markers, though in clusters 2 and 8 this was not as robust (Figure 3F).

2. Assigning glial cell type identities to clusters with known gene markers.

Briefly, according to our UMAP representations, the larval 7K data set clustered into 13 populations (Figure 2D) and adult 4.4K clustered into 10 populations (Figure 3D), where clusters are numbered from the largest (cluster 0) to the smallest. From these clusters, I isolated differentially expressed genes (log_fold-change (FC)<0.25, min.pct<0.1) and looked for unique combinations of at least 3 markers which, based on the known literature, could be used to provisionally assign cell type identities to each cluster. In addition, I also queried the specificity of new markers to strengthen the assignment of transcriptional profiles to cell types.

Using this method, we initially screened for our favorite glial populations: astrocytes, ensheathing glia, cortex glia and surface glia. Astrocytes were identified by the co-expression of GABA transporter (Gat) (Muthukumar, Stork, and Freeman 2014), astrocytic leucine-rich repeat molecule (alrm) (Doherty et al. 2009), Excitatory amino acid transporter 1 (Eaat1) (Stacey et al.

2010), wunen-2 (wun2) (Huang, Ng, and Jackson 2015) and ebony (e). Expression of these genes converged in larval clusters 5, 6, 11 and 12 (Figure 4A), and also in adult cluster 0 (Figure 5A). An even larger variety of astrocyte specific genes were found to be highly expressed in these clusters, confirming my findings (Figure 6A,B).

Ensheathing glia were identified by the co-expression of Excitatory amino acid transporter 2 (Eaat2) (Peco et al. 2016), Chloride channel-a (ClC-a) (Plazaola-Sasieta et al. 2019), axotactin (axo) (Davie et al. 2018) and CG9657 (Davie et al. 2018). Expression of these genes converged in larval cluster 3 (Figure 4B), as well as adult clusters 1 and 4 (Figure 5B). An even larger variety of astrocyte specific genes were found to be highly expressed in these clusters, confirming my findings (Figure 6C,D).

Cortex glia were identified by the co-expression of wrapper (Noordermeer et al. 1998), zydeco (zyd) (Melom and Littleton 2013) and akap200 (Davie et al. 2018). These genes showed high expression in larval clusters 0 and 2 (Figure 4C) and adult cluster 7 (Figure 5C). An even larger variety of astrocyte specific genes were found to be highly expressed in these clusters, confirming my findings (Figure 6E,F).

Lastly, surface glia were identified by co-expression of secreted protein, acidic, cysteine-rich (SPARC), moody and multi drug resistance 65 (Mdr65). These genes had the highest expression in larvae cluster 4 (Figure 4D) and adult cluster 3 (Figure 5D). An even larger variety of astrocyte specific genes were found to be highly expressed in these clusters, confirming my findings (Figure 6G,H).

Our glia-enriched data sets also contained several other non-neuronal cell types, perhaps because their transcriptional profiles are in some ways related to glia. To identify the remaining

clusters, I screened the data set for genes of known cell types in Drosophila – looking once again for combinations of genes with high expression (Figure 7). I identified different cell types in the two data sets. Specifically, in the larval data set (Figure 7A), cluster 0 co-expressed the progenitor cell marker SoxNeuro (SoxN) (Cremazy, Berta, and Girard 2000) and string (stg), a marker of a proliferative cell state (Edgar et al. 1994). Cluster 1 did not show any known combinations of gene markers. Clusters 7 and 9 co-expressed markers of neuroblasts (seven up (svp), high mobility group protein D (HmgD), modulo (mod) and suppressor of variegation 205 (Su(var)205)), however, cluster 9 also specifically had markers of cell cycle progression (geminin, cyclin A (cycA), proliferating cell nuclear antigen (PCNA), topoisomerase 2 (top2) and cyclin-dependent kinase 1 (Cdk1)) and of midline glia (Netrin A (NetA), Netrin B (NetB), slit (sli) and CG9336) suggesting that cluster 7 are neuroblasts and cluster 9 is midline glia potentially with undifferentiated precursors. Cluster 8, like cluster 1, did not show any known combinations of gene markers. Lastly, cluster 10 expressed markers of fat body (sestrin (Sesn), fasciclin 2 (Fas2), cystatin-like (Cys) and hexokinase A (Hex-A).

In the adult data set (Figure 7B), cluster 2 had a mixture of neuron and glial markers as noted earlier (Figure 3F). Cluster 5 did not show any known combinations of gene markers. Cluster 6 had expression of a combination of hemocyte markers (Drosomycin (Drs), Hemolectin (Hml), Serrate (Ser), Serpant (srp) and Papilin (Zelzer, Wappner, and Shilo). Cluster 8 had photoreceptor marker expression (elav, nSyb, no receptor potential A (norpA), transient receptor potential (trp), muscleblind (mbl), Arrestin 2 (Arr2) and neither inactivation nor afterpotential C (ninaC)). Cluster 9, like cluster 5, did not show any known combinations of gene markers.

To reveal novel genes that could define clusters, I obtained the top 10 genes in each cluster and plotted them against other clusters to see if there was specificity. I found that the most highly expressed gene markers in every cluster were fairly restricted to their respective cluster (Figure 7C,D). Together, this demonstrates that glial cell types of interest can be distinguished by combinatorial expression of a limited set of differentially expressed gene markers.

I was then curious to better understand the relationships of these cell populations and to see if I could start to corroborate knowledge from the literature. I used a dendrogram representation to analyze the hierarchical relationship of these clusters to one another. Looking at the larval data set, I noticed that astrocyte clusters 5 and 6 are related to ensheathing glia cluster 3 (Figure 4E). This was somewhat expected since Dr. Emilie Peco in our lab has previously shown that astrocytes and ensheathing glia in the larval ventral nerve cord are derived from the same progenitor cell (Davie et al. 2018). I also noticed that the larval glial clusters did not all branch off together, in fact, astrocyte clusters 5 and 6 branched from the same node and astrocyte clusters 11 and 12 from another – this points to a fundamental difference in these pairs of astrocyte clusters (discussed in more detail later). Seeing that clusters 7 and 9 had such similar expression of known markers it is not surprising to see that they branch together. Interestingly, the adult 4.4K data revealed two ensheathing glial clusters, consistent with a published report on distinct thoracic and neuropil ensheathing glia (Kremer et al. 2017), and is also consistent with the cluster identification of the Aerts lab – however, we have not assigned them to particular identities due to a lack of distinguishing marker genes. Most interesting from the investigation of the adult dendrogram (Figure 5E) was that astrocytes and ensheathing glia 1 but not 4 were branching from

the same node, suggesting that there could be a difference in origins of the two adult ensheathing glial clusters (discussed further later).

Of all the cells in the 7K data set, astrocytes, ensheathing glia and cortex glia together make up 42.9% of the entire data set, split almost evenly between each other (astrocytes: 14.4%, ensheathing glia: 13.8% and cortex glia: 14.7%) (Figure 4E). Of all the cells in the 4.4K data set, these cell types make up 63.7% of the entire dataset however, astrocytes (31%) and ensheathing glia (29%) make up a much larger proportion of that when compared to cortex glia (3.7%) (Figure 5E). There is no currently accepted proportion of glial cells for Drosophila larvae, however, for adults, astrocytes and ensheathing glia, have published estimates based on counting cells in situ (34% and 27%, respectively), which is consistent with what I have found. On the other hand, for cortex glia we identified much lower proportions than the published estimate (20%) (Kremer et al. 2017). It is possible that cortex glia might be poorly recovered in the scRNAseq workflow due to their lamelliform processes that elaborate around neuronal cell bodies (Pereanu, Shy, and Hartenstein 2005). Overall our findings demonstrate that clustering of scRNAseq data can be used to discriminate the major types of differentiated glia in both the larval and adult CNS.

3. Are some glial clusters restricted to distinct larval stages or subregions of the CNS?

The larval data from the Zlatic lab was collected at distinct stages of larval development: 1 hour (h) and 24h after larval hatching (ALH) corresponding to first instar, 48h ALH (second instar), and 96h ALH (third instar). In addition, the larval data was collected either from the entire CNS or from dissections that isolated the brain from the ventral nerve cord. All together the larval data came from 29 distinct samples, 5 of which were not used (imaginal disc, salivary gland or samples from 96H ALH that did not pass quality control parameters) (Figure 8A) each having a variable number

of larvae (range – 1 to 10) (Figure 8B). Cells in every major cluster were derived from a broad range of these samples (Figure 8C), suggesting the clustering was not strongly influenced by bias arising from differences in the quality of sequencing between samples.

With the larval data from the Zlatic lab, I was in a unique position to query whether the glial clusters we identified might be restricted to a particular larval stage or CNS region. To do this I mapped the information related to sample stage and region onto the UMAP representation of the clusters (Figure 8C), plotted the number and proportion of cells in these groups relative to the entire 7K data set (Figure 8E), and plotted the proportion cells in each cluster that was derived from a particular stage (Figure 8F) or CNS region (Figure 8G).

I noted that three clusters are composed solely from cells sampled at 1h ALH stage: two of which are astrocytic clusters 11 and 12 having few cells each (Figure 8E). Perhaps these represent cells in a state of immature astrocytic differentiation that does not persist, and so I chose to exclude these astrocyte clusters from further consideration, focusing on the major astrocyte clusters 5 and 6 instead.

It appears that the largest proportion of cells had a higher tendency for samples at 48 h ALH and that there is an underrepresentation of the 96h ALH stage (Figure 8D and F). This is apparent when looking at astrocytes, ensheathing glia and cortex glia, who have at most, 20% composition from stages other than 48h (Figure 8F). Where we see more distinction is in the tissue composition of these clusters (Figure 8G). I found that none of the major glial clusters could be found only in the brain or VNC, however the proportions of their compositions varied (Figure 8G). Astrocyte cluster 5 is predominantly derived from the brain whereas cluster 6 is predominantly derived from the CNS. Ensheathing glia appear to be relatively evenly distributed between samples

derived from CNS and brain, with a few VNC. On the other hand, cortex glia are almost solely from CNS but with no brain-only representation even though the brain samples make up almost 40 percent of all the samples. I noticed that the UMAP representation of this distribution in glial clusters was irregular in some instances, but in others it seemed that cells from particular stages/tissues were regrouped – overall, tendencies largely maintain representation of the sample proportions (stage distribution) in other clusters.

To explore this heterogeneity further, I needed to get more detailed information about the clustering within these populations. Therefore, I increased the resolution of the clustering analysis from 0.2 to 0.6 (Figure 9A). With increased resolution, astrocyte (clusters 5 and 6) formed two sub-clusters each (clusters 5a, 5b, 6a and 6b), ensheathing glia (cluster 3) formed 3 sub-clusters (3a, 3b and 3c) and cortex glia (cluster 2) formed two sub-clusters (2a and 2b) (Figure 9B).

Mapping the information related to larval stage and CNS region onto these sub-clusters, I found that astrocyte sub-cluster 6b, ensheathing glia sub-cluster 3a and cortex 2a were restricted to samples at 48h CNS (Figure 9C,D,F). Ensheathing glia sub-cluster 3b appears brain-specific since its cells came from whole CNS and brain samples, but not samples derived from VNC only (Figure 9D). Sub-clusters of astrocyte cluster 5 both showed greater proportions from the brain but remained quite diverse (Figure 9E). The remaining sub-clusters (6a, 3c and 2b) clearly showed a full representation of all tissues and stages (Figure 9C,D,F). Altogether, some but not all the observed cluster heterogeneity could be attributed to larval stage (astrocytes 6b, ensheathing glia 3a and cortex 2a) or CNS region (ensheathing glia 3b).

4. Using transcriptional profiles to obtain molecular signatures for the major classes of differentiated glial cells.

To better understand the molecular features that make clusters of differentiated glial cells distinct from one another, I extracted entire sets of DEGs from target clusters within the larval 7K and adult 4.4K data (logFC> 0.25, p_val_adj<0.05). For my thesis research, I will continue to focus on astrocytes, ensheathing glia and cortex glia.

In the larval 7K data, astrocyte clusters 5 and 6, ensheathing glia cluster 3, and cortex glia cluster 2 were composed of 215, 324, 143 and 319 DEG markers respectively (Figure 10A). In the adult 4.4K data, astrocyte cluster 0, ensheathing glia clusters 1 and 4, and cortex glia cluster 7 each had 210, 367, 274 and 382 DEG markers respectively (Figure 10A). To get an initial idea of the comparisons of glial cells within data sets, I chose to use larval astrocyte cluster 5 (Figure 10B) as a representative astrocyte cluster and adult ensheathing glia cluster 1 (Figure 10C) as a representative ensheathing glia cluster.

From the larval data, I found that each of these clusters had a largely unique profile of markers: astrocytes (cluster 5, 173/215 (80.5%) unique markers, ensheathing glia (88/143 (61.5%) unique), cortex (294/314, 92% unique) (Figure 10B). In terms of markers common to different glial cell types, I found that astrocyte cluster 5 shared only 17% of DEG markers with ensheathing glia and 3% with cortex glia. Ensheathing glia shared only 25% of their markers with astrocytes and 13% with cortex glia, whereas cortex glia shared only 2% of their markers with astrocytes and 6% with ensheathing glia.

I found that the profile of markers was similarly unique for glial clusters in adults (Figure 10C). Astrocytes (173/210, 82.4% unique), shared 17.6% of it DEGs, split between ensheathing glia (12.9%), cortex glia (3.3%) or both (1.4%). Ensheathing glia (195/367, 53% unique) shared 47% of

its DEGs, split between astrocytes (7.4%), cortex glia (38.7%) or both (0.8%). Cortex glia (230/382, 60.2% unique) shared 39.8% of its DEGs, split between astrocytes (1.8%), ensheathing glia (37.2%) or both (0.8%). Evidently, ensheathing glia in both larval and adult drosophila share the largest proportion of their DEGs with other clusters.

This initial comparison of glial cells within each stage set the scene to next explore the DEG markers, along with more precise identities, that distinguished astrocytes, ensheathing glia and cortex glia between the two stages (larvae versus adults). I started by comparing all of the DEGs (all cell types) between the two data sets to get an idea of their similarity and found that there was an important overlap (1593 DEGs) in the markers defining the cells contained in drosophila larvae and adults – representing more than half of the DEGs in the larval data set and just under half of the DEGs in the adult data set (Figure 11A).

1.1. Astrocytes:

I compared the entire list of DEGs from astrocytes in larvae (cluster 5 -215 DEGs or cluster 6 – 324 DEGs) with those in adults (cluster 0 – 210 DEGs) and found that there were 93 genes overlapped between cluster 5 (Figure 11B) and adults and 70 genes overlapped between cluster 6 and adults (Figure 11C). I then isolated the top 10 DEG markers (ordered by avg_logFC) that were unique to larva, or unique to adults, and also those that were shared. Of those shared between cluster 5 and adults, and those shared between cluster 6 and adults, 7 were duplicated (bold in Figure 11B,C). When checked against all DEG markers for all clusters in the larval 7K and adult 4.4K data sets, only wun2, Gat and CG9394 stood out as a trio of markers seemingly forming a unique signature for Drosophila astrocytes in both larvae and adults. Interestingly, even though I am not including

them in this analysis, the small larval astrocyte clusters (11 and 12) also expressed this trio of markers, reinforcing these genes' specificity.

1.2. Ensheathing Glia:

Comparing lists of DEGs derived from ensheathing glia from the larval data set (cluster 3 - 143 DEGs), to those in adults (cluster 1 - 367 DEGs, cluster 4 - 274 DEGs), I found 80 genes in common between larvae and adult cluster 1 (Figure 11D) and 57 genes in common between larvae and adult cluster 4 (Figure 11E). I isolated the top 10 DEGs (ordered by avg_logFC) that were unique to larva, unique to adults and those that were shared between the two and of the 10 DEGs in common between cluster 1 and larvae and those in common between cluster 4 and larvae: 8 were duplicated (bold in Figure 11D,E). When checked against all DEG markers for all clusters in the larval 7K and adult 4.4K data sets, only Elal and DNasell were uniquely expressed in the larval and adult ensheathing glia clusters - seemingly forming a unique signature for Drosophila astrocytes in both larvae and adults.

1.3. Cortex Glia:

Both larval and adult data sets had only one cortex cluster – comparing markers in larvae (cluster 2 - 319 DEGs) to those in adults (cluster 7 - 382 DEGs), I found they had 103 genes in common (Figure 11F). Focusing on the top 10 genes (ordered by avg_logFC) that were unique to larvae, unique to adults and those that were common to both clusters - only CG40470 and slc45-1 were uniquely expressed in the cortex glia clusters of both data sets when checked against the DEG markers for all clusters in the larval 7K and adult 4.4 data sets - seemingly forming a unique signature for Drosophila cortex glia.

5. Use of Gene Ontology to characterize glial clusters

To explore molecular and functional specializations of different glial cell types, we used approaches involving Gene Ontology (GO) terms to analyze lists of DEG markers (adjusted p-value>0.05; log2(FC)>0.25). The GO bioinformatics initiative unifies the representation of genes to capture knowledge on gene function with a controlled vocabulary applicable to all organisms (Ashburner et al. 2000). There are three overarching themes of gene ontology, we will focus on two: biological process (BP) - the biological objective to which the gene or gene product contributes and molecular function (MF) - the biochemical activity (including specific binding to ligands or structures) of a gene product. To study gene ontology, we used Cytoscape, an open source software platform for visualizing complex networks of genes (Shannon et al. 2003). Within Cytoscape, the ClueGO plug-in can decipher biological networks specifically for scRNAseq data, allowing for a visualization of terms and comparisons of functional annotations. The resulting lists of terms (ranging in number from 5 to 134 BP and 5 to 34 MF) were curated to only show the 20 most enriched and relevant terms.

As proof of principle, I found that all glial clusters exhibit enrichment in terms related to the nervous system: from its development with terms like gliogenesis and neuron formation, to functions such as learning or memory (Figures 12-14). Other enriched terms confirmed typical functions of glial cells including metabolism, ion channels and the transport of molecules across membranes, like amino acids and neurotransmitters.

5.1 Astrocytes

Looking at non-redundant terms, we found that GO analysis with marker genes from larval astrocyte cluster 6 returned a highly significant enrichment in protein synthesis processes, from ribosomal subunits biogenesis and ribosome assembly, to translation initiation and elongation (arrows in Figure 12A). This prominent enrichment was not present in either the other major larval astrocyte (cluster 5) (Figure 12B) or the adult astrocyte clusters (Figure 12C). This was correlated with enrichment in the molecular functions for RNA binding, translation regular activity and other translation related mechanisms (arrows Figure 15A). Cluster 6 also differentiated from cluster 5 and adult astrocytes due to a lack of transmembrane transporter activities (Figure 15A,B) and a response to ethanol (Figure 12). These results suggest that a great pool of larval astrocytes (cluster 6) is in a different cellular state where protein synthesis is very active.

Seemingly unique to the larval data sets, were processes regarding cell junctions, potentially important during this time in development if cells are still forming adhesions and establishing communication with neighbouring cells (Figure 12). In conjunction with the processes described earlier common to all glial cells, all of the astrocyte clusters had enrichment in terms regarding cell communication, chemical synaptic transmission as well as various metabolic processes (glutathione, ribonucleoside diphosphate and amino acids), which agrees with their well-known functions mentioned previously (Figure 12).

5.2 Ensheathing Glia

Looking at larval ensheathing glia, the analysis showed no transmembrane transport activities and also lacked neural processes – noteworthy is that this could be due to very few enriched terms, all of which were related to metabolism with a particular emphasis on carbohydrates (Figure 13A).

As mentioned earlier, in the adult dataset we identified 2 ensheathing glia clusters (cluster 1 and 4). Our GO analysis revealed that these 2 clusters are enriched in rather similar biological processes and molecular functions including oxidoreductase activity, homeostasis, transmembrane transport and diverse metabolic processes (Figure 13B,C). Compared to the other adult cluster, ensheathing glia cluster 4 showed a greater enrichment in transmembrane transport and lacked processes directly related to the nervous system (Figure 13C). This supports the idea already in the literature of two co-existing ensheathing glia populations with specific functions (Kremer et al. 2017). However, a deeper analysis is required here to uncover exactly how these populations differ.

5.3 Cortex Glia

Cortex glia clusters in both larvae and adult showed very similar GO profiles including pronounced enrichment for terms related to cell communication (Figure 14), not surprisingly given their position in close proximity to neuronal cell bodies (Pereanu, Shy, and Hartenstein 2005; Freeman 2015; Awasaki et al. 2008; Kremer et al. 2017) and to other glial cells such as the surface glia (Kremer et al. 2017). I also found that they expressed enriched MF typical to glia such as transmembrane transporter activities (Figure 15C,D) and enriched biological process involving nervous system function, growth, cell differentiation/fate commitment, metabolic homeostasis (response to stimuli and response to nutrients), behavior (Figure 14). Interestingly, we found specific enrichment for lipid-related processes (lipid localization, lipid transport or response to lipid), which correlates with evidence in the literature, that cortex glia accumulate lipid droplets (Kis et al. 2015).

6. Comparing Drosophila glial cell types with glial cells in the mouse brain.

To increase the breadth of this research, we looked to see how the transcriptional profiles of glia in flies and mammals compare, taking advantage of the mouse brain atlas from the McCarroll lab (Saunders et al. 2018). Obtained with high-throughput single-cell RNA sequencing; the data set consists of 690,000 cells that were obtained from nine regions of the adult mouse brain (frontal cortex, striatum, globus pallidus externus & nucleus basalis, entopeduncular nucleus & subthalamic nucleus, substantia nigra & ventral tegmental area, cerebellum, posterior cortex, thalamus and hippocampus). This data set contains twelve mammalian cell types: **neurons,** **astrocytes**, **oligodendrocytes**, **microglia/macrophages**, **polydendrocytes** (oligodendrocyte progenitor cells), components of the blood-brain-barrier (or vasculature cell classes): endothelial cells (**tip** and **stalk**) and **mural cells**, cells native to the ventricles or **ependymal cells**, and finally cell classes undergoing **neurogenesis** and **mitosis** (Figure 16A).

We wondered if glial cell types in Drosophila had specific counterparts in mice that could be identified by expression of a specific set of orthologous genes. For example, might astrocytes in both species be more closely related than they are to other glial cells. Might cortex glia also show molecular similarity to mouse astrocytes; or ensheathing glia to oligodendrocytes (Pereanu, Shy, and Hartenstein 2005)?

The mouse data from the McCarroll lab was not processed for glia-enrichment, and for this reason we returned to the original, non-preprocessed Drosophila data: larval 100K and adult 57K from the Zlatic and Aerts labs respectively. Clustering of the larval 100K data set resulted in two astrocyte clusters (cluster 38 and 47), one ensheathing glial cluster (cluster 31) and two cortex glia clusters (cluster 20 and 23). Clustering of the adult 57K data set resulted in one astrocyte cluster

(cluster 10), two ensheathing glial clusters (cluster 14 and 35) and one cortex glia cluster (cluster 60). With the lone exception of finding two cortex glia clusters in larvae – these clusters are consistent with our clusters from the larval 7K and adult 4.4K data sets.

For each of the clusters from both data sets, I extracted the top 200 DEG markers and determined their mouse orthologs using DIOPT (DRSC Integrative Ortholog Prediction Tool) (Hu et al. 2011). Our collaborator Dr. Todd Farmer next ran statistical tests in R using Seurat 3.6.3 to identify orthologs that were DEGs (min.pct of 0.1 and a logfc.threshold of 0.25) in the McCarroll mouse brain scRNAseq data. These DEG marker orthologs were then plotted onto heatmaps of SCTransformed scaled data using a subsample of 500 mouse brain cells (Figure 16B,D).

We found that a small proportion of Drosophila glial markers were found in mouse brain subtypes - from an original 200 markers, only between 2 and 19 of these were differentially expressed mouse orthologs that could be found in a mouse brain cell type (Figure 16C,E). On first glance, there is a similar pattern of expression for all Drosophila cell types onto the mammalian cell types regardless of the data set. Based on the heatmap overlap, Drosophila glia resemble mammalian glia more so than neurons but don't have specific sub-type similarity (Figure 16B,D).

From the larval 100K data, astrocyte cluster 38 and 47 had different expression patterns: cluster 38 showed high gene overlap with most mammalian cell types: astrocyte, oligodendrocytes, ependymal cells, polydendrocytes, mitotic cell, neurogenesis, mural and endothelial cells. On the other hand, cluster 47 had high expression in only four cell types: oligodendrocytes, ependymal and endothelial cells (Figure 16B,C). This suggests that astrocyte cluster 47 has a transcriptome that is more specifically correlated to mouse cell types. Ensheathing glia on the other hand showed high overlap (8 genes) onto astrocytes, oligodendrocytes and

ependymal cells (Figure 16C). Drosophila cortex glia cluster 20 showed more robust overlaps to mouse cell types then cluster 23 but they commonly showed overlaps in ependymal cells, mitotic cell and endothelial stalk cells (Figure 16C). We found that in the adult 57K that all three Drosophila cell types showed very high gene expression overlap (≥ 9 genes) in mammalian astrocytes, oligodendrocytes and endothelial cells and showed low gene expression overlap (≤ 4 genes) in mammalian neurons (Figure 16D,E). We also observed that ensheathing glia 14 has almost double the number of genes in the progenitor cell types when compared to the ensheathing glia cluster 35 (Figure 16E).

Based on the number of genes which overlap between the drosophila and mammalian data sets, the drosophila larvae have many less genes in common with mice when compared to the adults (Figure 16C,E). However, as established by the heatmap patterns of expression, there does not seem to be a correlation between specific cell sub-types (Figure 16). To better appreciate the overlap of genes, we merged the fly glial lists where there was more than one cluster for a single cell type. Doing this did not impact the findings in terms of the number of overlapping genes and we still did not see cell-specific overlaps between species. Seeing that the numbers of orthologs did not establish a clear relationship between specific types of glia in flies and mice, I wanted to look at comparisons of specific genes instead. One of the first things I noticed doing this, was the expression of high numbers of ribosomal proteins (RPs) (Figure 17).

7. Ribosomal Proteins are highly conserved in glial cells between flies and mice

Ribosomal proteins are involved in assembly of either small 40S ribosomal subunits (Rps) or large 60S ribosomal subunits (Rpl) (Zhou et al. 2015). RnrS (RRM2, ribonucleotide reductase) shows significant expression only in mitotic (very small cluster) and neurogenic mouse subtypes (Figure

17A). Interestingly, it has been found to be key for human glioblastoma proliferation (Li et al. 2018) and in our larval 7K dataset, it is expressed in our neuroblast cluster (cluster 9). I found that the enrichment of these RPs significantly skewed our analysis (Figure 18). For example, astrocytes showed a high ribosome content (30-77%) in a rather non-specific way in most mouse subtypes, but are not well expressed in mouse astrocytes, neurons and mitotic cells (Figure 18C). Drosophila ensheathing glia map onto mouse cells with lower ribosomal content (10-50%) (Figure 18C) and always the same four proteins: Rpl13a, Rplp1, Rplp22 and Rplp2 (Figure 17). Least affected by ribosomal proteins are the cortex glia, who map onto the mouse data set with low ribosomal content (0.1-20%) (Figure 18C) and by a single protein: Rps11, which was uniquely found in this glial subtype (Figure 17A).

8. Exploring genes in common to larvae and adult Drosophila with ortholog overlaps in mice

Literature suggests that mouse astrocytes and oligodendrocytes would be the closest equivalents to Drosophila neuropil glia (Yildirim et al. 2019). For this reason, we focus on the overlaps onto these cell types even though we demonstrated no correlation with the numbers of genes mapping onto mouse cell types - extrapolating these genes to direct our lab into a direction with a focus on material most interesting for *in vivo* analyses. Due to their non-specific enrichment, I removed all RPs from the analysis to uncover underlying genes that were more specifically distributed (Figure 19). Based on the number of genes (not including RPs) we found that Drosophila Glia subtypes are more closely related to mouse macroglia and vasculature subtypes.

Based on numbers, larval astrocytes seem more closely related to mouse oligodendrocytes, astrocytes, ependymal cells and cells of the blood vessels (Figure 18A), whereas

adult astrocytes seem more specifically related to astrocytes and oligodendrocytes (Figure 18B). Looking at specific genes from both larval and adult Drosophila astrocytes, I found a few with orthologs that mapped onto mouse astrocytes: dally, Eaat1, Gat and Gs2 (Figure 19A), which have previously been established as being astrocytic markers. Of these 4 genes, dally was the only gene expressed uniquely in mouse astrocytes. Drosophila astrocytes also showed orthologs that overlapped onto mouse oligodendrocytes: CAH1, CHA7 and Gs2 – CAH1 and CAH7 both being expressed uniquely in oligodendrocytes (Figure 19B).

Larval and adult ensheathing glia converge onto astrocytes, oligodendrocytes and ependymal cells but adult ensheathing glia additionally have expression in others mammalian cell types like polydendrocytes, and blood vessels cells (Figure 18). In both larvae and adult drosophila ensheathing glia, Gs2, Tsp42Ee had orthologs that mapped onto mouse astrocytes (Figure 19C), amongst others. They both also showed overlap of CAH1, Gs2, Gel and Tsp42Ee orthologs onto mouse oligodendrocytes (Figure 19D) – CAH1 solely showing unique expression in oligodendrocytes. Interestingly, the tetraspanin superfamily can be viewed as 'molecular facilitators' of diverse cellular functions, from cell adhesion to signal transduction (Maecker, Todd, and Levy 1997), in particular, Tsp42Ee has been previously shown to regulate septate junctions in Drosophila midgut (Izumi et al. 2016) pointing to a conserved processes between drosophila ensheathing glia and mammalian glia.

As for cortex glia: larvae seem more closely related to endothelial stalk cells, ependymal cells, astrocytes, oligodendrocytes and mitotic cells (Figure 18A) whereas adults are more highly related to astrocytes and oligodendrocytes, followed by stalk cells, ependymal and macrophages (Figure 18B). Cortex glia orthologs that were common to both larvae and adult drosophila were

more numerous than for astrocytes and ensheathing glia (Figure 19). Those that mapped onto mouse astrocytes included: Fabp, CG1764, Irk3, CD98hc, aay and Acbp2 (Figure 19E), where CG1764 was the only gene with unique orthologous expression onto mammalian astrocytes. They also showed overlap onto mouse oligodendrocytes from orthologs of: Fer1HCH, Acbp2 and Irk3 (Figure 19F), amongst others.

Overall, there were not many genes derived from both larval and adult data sets that showed orthologous mapping onto mammalian astrocytes and/or oligodendrocytes. Generally speaking, it was rare to see any drosophila gene orthologues with unique expression onto a singular mammalian cell type (regardless of the cell type). Interestingly, however I noticed that DIP-B, expressed in adult drosophila cortex glia, which encodes a peptidase involved in protein metabolism was uniquely found in mammalian astrocytes and oligodendrocytes (Figure 19E,F). I also found that CG6287 was the only gene represented in adult astrocytes, ensheathing glia and cortex glia and additionally had expression uniquely in mammalian astrocytes and oligodendrocytes (Figure 19). These genes would be interesting to prioritize in future studies

9. The role of transporters in Drosophila glia.

A major role of glia is to regulate brain physiology, which is largely done by regulating the movement and concentration of substances in and out of the extracellular fluid – a role orchestrated by transporters (Hediger et al. 2004). The largest and most highly conserved class of transporters are solute carrier (SLC) proteins (Ren, Chen, and Paulsen 2007). Not only are they highly conserved between species, they are also abundant in glia and their function in the brain remains relatively poorly understood. Seeing how very little is known about how most SLC transporters might regulate complex brain functions and/or behavior via selective movement of

materials, and the fact that they can be relatively easily knocked out or modified in genetic model organisms (such as *Drosophila melanogaster*) (Featherstone 2011), puts our lab in a unique position to study them further.

I found SLC transporters as DEG markers in Drosophila astrocytes, ensheathing glia, and cortex glia, and so I decided to investigate this important family of proteins in greater detail in these cell types. I also investigated this in surface glia seeing as they are gate keepers - controlling the influx/efflux of substances between the hemolymph and the CNS of Drosophila. From GO term analysis we isolated all the SLC transporter genes that were DEG markers for these glial clusters in the larval 7K and adult 4.4K data sets. The adult data set returned many more SLC genes (58) than the larval data set (25) and included 80% (20/25) of the same larval SLC's (Figures 20 & 21). Astrocytes showed the lowest number of SLC gene expression when compared to the other 3 cell types in both data sets (Figures 20 & 21). Larval astrocyte clusters 5 and 6 had near identical expression of SLC's (Figure 20), similarly, when I compared the two adult ensheathing glia clusters 1 and 4, I found they had very similar expression patterns (Figure 21). In adults, but not larvae, surface glia had a high number of largely distinct SLC transporters - differentiating them from all other cell types (Figure 21). Cortex glia of both larvae and adults also showed a unique pattern of SLC expression, with some overlap with ensheathing glia compared to other cell types. Overall, astrocytes, ensheathing glia, cortex glia and surface glia have distinct patterns of highly expressed SLCs which likely contribute to their specialized roles in regulating solutes such as ions and neurotransmitters, thereby influencing CNS function and behavior.

Discussion

Glial cell heterogeneity poses a daunting challenge for research programs aimed at studying how glia contribute to health and disease. Although it is not new that glial cells show intercellular diversity, our understanding is limited of how these cells are heterogeneous with respect to their marker expression and how that might translate into functional specializations. In my thesis research, I first aimed to distinguish glial cells via cluster analysis on scRNA-seq data from neural tissues at larval and adult stages. I have confirmed that glial cells can be distinguished at the level of the transcriptome, and have, for the first time, used glia-enriched scRNAseq data sets to uncover more precisely how these glial subtypes compare in the larval and adult CNS of Drosophila.

With data sets enriched in glial cells, I identified known cell types in Drosophila larvae and adults and analyzed their molecular relationships to one another. Glia enrichment was an important step for this analysis because I was looking to find differences between glial subtypes and did not want factors differentiating glia from non-glia cell types to have undue influence on my findings. As expected, both the larval and adult glia enriched data sets had strikingly more glial markers than neuronal markers, but in the larval 7K data set the expression levels of recognized glial markers seemed relatively lower than in the adult 4.4K data set (Figure 2F), despite the fact that many Drosophila glial cell types are fully differentiated in the larval stage. It's unclear whether this reflects real differences in glial marker expression levels, or results from data processing and normalization of the distinct data sets.

I next set out to explore the molecular relationship between glial cell types from distinct stages of life and different regions of the CNS to define stage-specific specializations and

investigate regional heterogeneity. Molecular signatures can be used to differentiate distinct cell populations from one another, and in general my research newly identifies lists of genes whose expression characterizes astrocytes, ensheathing glia and cortex glia. We found it rare that genes were uniquely expressed in a single cell type and not in others, though I did nonetheless identify a handful of unique cell-type specific markers with high expression in both larvae and adults. In my opinion, these markers ought to be prioritized for future investigations as they could provide insight into functional specialization of these different cells.

Based on the overlap of markers between clusters, astrocytes were seemingly the most distinct, sharing few markers with ensheathing glia and cortex glia. In contrast, ensheathing glia seemed the least distinct, sharing 38% (in larvae) and 47% (in adult) of their markers with astrocytes and cortex glia. Similar results were observed in adults. Perhaps astrocytes and cortex have more specialized functional roles in the CNS of Drosophila, whereas ensheathing glia might serve more generalized functions. This possibility is supported by functional enrichment using Gene Ontology, where ensheathing glia had the lowest number and smallest variety of enriched terms.

Astrocytes have long been regarded as a homogenous cell population, with some suggestion that they exhibit molecular heterogeneity, possibly to match astrocyte functions to the local needs of neurons nearby (Chaboub and Deneen 2012; Zhang and Barres 2010). I show this quite convincingly in my thesis, where clustering of larval drosophila resulted in 4 astrocyte populations, two of which (clusters 11 and 12) had minimal number of cells (53 and 40 respectively) and were derived solely from tissues at 1H ALH. It is possible they represent transient or immature populations, but these “young astrocytes” were removed from further consideration

due to their rarity. Focusing on the two main clusters (5 and 6), I found that DEG markers of cluster 5 had a greater overlap with adult astrocytes (93/215 – 43%) than did the DEG markers of cluster 6 (70/324 – 21.6%). Of the markers that were common, 3 were unique to astrocyte clusters: CG9394, Gat and wun2 – of which only CG9394 has not yet been characterized as being astrocyte-specific. Cluster 5 was also distinct from cluster 6 because it was more specific to the larval brain (as opposed to VNC of whole CNS). Further differentiating these astrocytes, the GO analysis suggested that cells of cluster 6 are specialized towards a heightened state of protein translation. Interestingly, I found that among known functions, both larval and adult astrocytes shared a novel enriched function related to response to ethanol - studies in Drosophila have shown surface glia play a role in ethanol tolerance but to date this role has not been attributed to astrocytes (Bainton et al. 2005; Parkhurst et al. 2018). We know that mice show astrocyte regional specificity in their spinal cord and in the cortex (Molofsky et al. 2014; Bayraktar et al. 2020), which would be interesting to investigate further in flies. My current analysis lacks the resolution to address this, however by sub-setting our astrocyte clusters in larvae and adult, and re-clustering them, we could study their molecular composition more precisely to see if we can uncover further specializations. This followed by in situ hybridization would shed light on regional specializations that could confer added cortical organizational complexity not yet known.

For ensheathing glia, others have identified two distinct populations (neuropil and tract) based on morphology (Kremer et al. 2017). Our findings have corroborated this in adults, where we found one larval ensheathing glia population and two adult ensheathing glial populations. Ensheathing glia in the adult CNS are capable of functioning as phagocytes, taking up and digesting cellular debris from injured neurons (Sonnenfeld and Jacobs 1995). Perhaps related to this, one of

the unique markers we found for ensheathing glia is DNasell – localized in lysosomes, and essential for degrading nuclear DNA of engulfed apoptotic cells (http://flybase.org/reports/FBgn0000477). Another unique DEG marker of ensheathing glia was Elal, which encodes Elastin-like, a potential component of extracellular matrix. I found that both adult clusters of ensheathing glia (clusters 1 and 4) had high expression of genes related to metabolism, homeostasis and cellular responses to stress. It however remains unknown how these clusters might relate to the morphological classes of tract and neuropil ensheathing glia. From the dendrogram of the larval data set, astrocytes and ensheathing glia are tightly related, which is consistent with our demonstration that they share a common lineage in larvae. Interestingly, through shared markers larval ensheathing glia resemble adult ensheathing glia cluster 1 more than they do cluster 4 which in the dendrogram appears more closely related to neuroblasts and midline glia.

To complete my thesis project, I considered the functional relevance of our findings via gene ontology analysis and examined the conservation of molecular features between glial cells in Drosophila and mice. From the Gene Ontology analyses, we found that the larval ensheathing glia were enriched for a potential role in carbohydrate metabolism. Interestingly, genes required for carbohydrate metabolism are usually found in the fat body, intestine, and Malpighian tubules (Havula et al. 2013), once again suggesting a common feature between ensheathing glia and fat body. Carbohydrate metabolism involves the breakdown of carbohydrates into glucose, which is then stored as glycogen – we have seen that aberrant accumulation of glycogen in neurons coincides with aggressive neurodegeneration in humans (Lafora disease) (Gentry et al. 2018) and has been shown, in Drosophila, to lead to neuron loss, reduced locomotion, and shortened life span (Duran et al. 2012). Given that ensheathing glia form close associations with newly-born

neurons in the larval brain (Dumstrei, Wang, and Hartenstein 2003), along with their enrichment for carbohydrate metabolism processes, this could indicate a potential role for ensheathing glia in neuroprotection.

Of the three glial subtypes I focused on for my thesis, cortex glia have been the least extensively described in the literature. My analysis of larval and adult DEGs in cortex glia found unique expression of two genes previously undescribed as cortex markers: CG40470 and Slc45-1. Cortex glia exhibit activity-dependent calcium oscillations, are able to regulate seizure susceptibility (Melom and Littleton 2013), and have been suggested to provide trophic support to neurons (Stork, Bernardos, and Freeman 2012). More recently, there has been evidence suggesting that cortex glia are able to form lipid droplets under conditions of hypoxia and oxidative stress, which could in turn protect neural stem cells from oxidative damage (Kis et al. 2015; Bailey et al. 2015). We know that oxidative stress can fuel lipogenesis in neurons, and due to their close contact and communication, these lipids could be transported and stored in glia as lipid droplets to prevent neurodegeneration (Liu et al. 2017; Liu et al. 2015). My analysis has shown biological processes and molecular functions in larvae and adults with elevated representation of lipid localization, lipid binding, lipid transport, and responses to lipids, which strongly reinforces this idea that cortex glia would have a specific role in preventing the accumulation of lipids in neurons. Dendrograms of hierarchical clustering also demonstrated that cortex glia are closely related to fat body, which could relate to their functions with lipids (Colombani et al. 2003). Together, this suggests cortex glia are an interesting target for future studies investigating neurodegenerative diseases caused by accumulation of lipids in neurons.

Evolutionary conservation of glial cells:

We looked to see how the transcriptional profiles of glia in flies (larval 100K and adult 57K) compare to those in mammals by taking advantage of the mouse brain atlas from the McCarroll lab (Saunders et al. 2018). It has been suggested in the literature that there might be Drosophila glial cell types with specific counterparts in mice (Pereanu, Shy, and Hartenstein 2005). We explored this idea and found that Drosophila glia are related to mouse glia more closely than to neurons at a transcriptomic level.

I found that the larval 100K data set was composed of two cortex glia clusters compared to the single cluster from larval 7K data set. I suspect that since many more cell types are present in the 100K data set, the principal components used to clusters cells were different and might have caused this clustering to change in cortex glia. Another possibility could be that one of these clusters is not truly a cortex glial cluster – presence of SoxN in cluster 20 could suggest a progenitor state. My analysis of specific genes orthologs which overlapped between species exposed a high expression of ribosomal proteins in all subtypes of both larval and adult Drosophila, particularly from astrocytes. To focus on other glia-specific genes I removed these ribosomal proteins from the analysis, revealing a small proportion of Drosophila glial markers that had orthologs expressed in mouse brain glial cell subtypes. From these, there appeared no obvious one-to-one molecular relationship between specific Drosophila and mouse glial cell types. On one hand this could suggests that glial subtypes in each species have become molecularly specialized and have not been broadly constrained through evolution to a particular molecular profile. However unlikely, it also raises the possibility of evolutionary convergence of some generalized glial features and functions onto cells that are otherwise not conserved. Our data does not clearly address this,

though further investigation of this interesting question about the origins and functions of glia cells is warranted.

Genes that did have expression in both larval and adult Drosophila glia, and whose orthologs are expressed in mammalian glia, are most interesting since they could illuminate ancestral, evolutionarily conserved roles for glial cells. I observed that astrocytes in Drosophila and mice share processes in transmembrane transport, and that Drosophila astrocytes and mouse oligodendrocytes shared metabolic processes. Of these genes, Gs2 was found in both larvae and adult astrocytes, with orthologous expression in both mammalian astrocytes and oligodendrocytes. Gs2 encodes a glutamine synthetase involved in both cellular metabolism and neurotransmitter (glutamate) processing. This suggests that its role in these processes is highly conserved. Fly ensheathing glia did not show clear functional relation to mouse astrocytes but did show that certain metabolic processes (i.e. carbonic anhydrase, cellular amino acid, glutamate and glucose metabolism) could be shared with mouse oligodendrocytes. Although cortex glia had the largest number of genes with orthologues that mapped onto mammalian astrocytes, there was also greater variety – many of which were related to expected glial functions such as metabolism and transport. Perhaps interesting for future studies, cortex glia shared with mouse oligodendrocytes expression of orthologs of an iron storage related protein.

Based on the number of overlapping orthologous gene markers with mice glia, Drosophila larvae glia have fewer in common than do Drosophila adult glia. I speculate this could reflect the possibility that the adult and larval CNS have different specialized requirements and that those of the Drosophila adult CNS may be more akin to those of the mature mouse brain. Because of this I identified genes derived from adult flies which mapped onto adult mouse cell types and found

that CG6287 was found in all three adult fly cell types, and uniquely mapped onto both mouse astrocytes and oligodendrocytes. CG6287 is an enzyme that is, to date, not well studied (https://flybase.org/reports/FBgn0032350), but is suggested to be involved in glycolytic catabolism and lipid biosynthesis (Lissemore et al. 1990). Likewise, another gene marker from adult fly glia whose ortholog was found uniquely expressed in mouse astrocytes and oligodendrocytes was Dip-B, which encodes a peptidase with function related to metabolism, reinforcing that a strongly conserved function of glia between these species is metabolic.

From the gene ontology analysis, we focused added attention on SLC transporters due to their functional specificity in transporting certain substances across membranes - regulating their movement between cells or intracellular compartments. I found astrocytes, ensheathing glia, cortex glia and surface glia have distinct but partially overlapping arrays of SLC transporters. In addition, I found similar expression of SLC transporters between the two larval astrocyte clusters 5 and 6, and likewise between the two adult ensheathing glial clusters 1 and 4. This supports the idea that glial cell types are specialized to support specific transmembrane transport activities. Showing the largest number of distinct transporters were the adult surface glia – most likely due to their position between the hemolymph and CNS requiring them to act as gate keepers for the specific substances being exchanged. Interestingly, I did not find this to be the case for larval surface glia, potentially because other cells support this role to a greater degree in larvae. Overall, there is a pattern of SLC transporters found for each glial subtype – some of which are shared between cell types as well as across stages – that can be explored further to better characterize the functions of Drosophila glia.

Future Directions

Until now, we have been able to deeply characterize three subtypes of Drosophila glia (astrocytes, ensheathing glia, and cortex glia), however we have identified several others worthy of added attention. I would particularly be interested in doing a deeper characterization of the transcriptome of surface glia, since it is essential that the interface between the CNS and the peripheral circulatory system be an effective regulator of ion balance, a facilitator of nutrient transport, and a barrier to potentially harmful molecules (Hawkins and Davis 2005). My clustering resulted in one population of surface glia and I have, in preliminary work not shown here, found that the top DEGs of surface glia expressed enrichment in processes related to transmembrane transport, ATP metabolic processes, cell junction organization and assembly as well as in the generation of metabolites – which supports previously described functions. I would like to further investigate a comparison with mammalian glia as well as explore specific molecular markers more deeply by comparing larvae to adults as well as identifying similarities and differences in the markers of astrocytes, ensheathing glia and cortex glia. For my thesis, I largely focused on the similarities found between glial clusters and between data sets, however it would be interesting to do an analysis on the DEGs and processes that differentiate these glial cell types to uncover more on if/which functions specialize during development.

Now that we have characterized glial cells and analyzed orthologs in mammals as well as done GO analysis, we can extrapolate that information to identify novel genes involved in crucial pathways in Drosophila with conserved functions in mice. Our lab is currently exploring this possibility, curating lists from these results in Drosophila and mice in order to apply them to *in vivo* genetic studies in flies. Work has already begun for a loss-of-function RNA interference screen for

factors expressed in glia that profoundly influence larval or adult locomotion, and sleep in adults (collaboration with E. Cho, Ph.D. and E. Peco, Ph.D). Such experiments have the potential of corroborating our in-silico findings and could uncover novel functions for glial cells in regulating CNS function and animal behavior.

Conclusion

My MSc thesis research project used data sets obtained through single cell RNA sequencing to confirm previous findings about molecular heterogeneity in glial cells of the *Drosophila melanogaster.* Using unsupervised clustering I was able to determine groups of cells based on similar transcriptional signatures within a sample and cells within a cluster were collectively labeled based on the average expression levels of canonical markers. I have found similarities and differences which can be used to isolate, identify and deeply characterize cell types based on gene markers, age and stage of tissue as well as molecular and biological functions. I have demonstrated how certain glial cells might evolve from larvae to adult and how orthologs of gene markers in fly glial cells compare to markers in mammalian glia. The results demonstrate how *Drosophila melanogaster* can be used to provide new insights on glial functions and their roles in the pathology of neurodegenerative diseases through the detailed characterization of glial profiles which revealed molecular features that may explain their distinct attributes and functions.

My *in-silico* findings will increase the breadth of understanding of glial cells in this important model system and has identified candidate genes for further analysis to understand what roles these distinct glial subtypes play in the proper function of the central nervous system. Importantly, my work will facilitate and provide fuel for future studies that will continue to expand our knowledge in the growing field of glial biology by serving as a resource to understand glial cell-type identity.

References

Abbott, N. J., L. Ronnback, and E. Hansson. 2006. 'Astrocyte-endothelial interactions at the blood-brain barrier', *Nat Rev Neurosci*, 7: 41-53.

Acosta, C., H. D. Anderson, and C. M. Anderson. 2017. 'Astrocyte dysfunction in Alzheimer disease', *Journal of Neuroscience Research*, 95: 2430-47.

Aguayo, A. J., J. Epps, L. Charron, and G. M. Bray. 1976. 'Multipotentiality of Schwann cells in cross-anastomosed and grafted myelinated and unmyelinated nerves: quantitative microscopy and radioautography', *Brain Res*, 104: 1-20.

Ahrens, M. B., M. B. Orger, D. N. Robson, J. M. Li, and P. J. Keller. 2013. 'Whole-brain functional imaging at cellular resolution using light-sheet microscopy', *Nat Methods*, 10: 413-20.

Allen, A. M., M. C. Neville, S. Birtles, V. Croset, C. D. Treiber, S. Waddell, and S. F. Goodwin. 2020. 'A single-cell transcriptomic atlas of the adult Drosophila ventral nerve cord', *Elife*, 9.

Allen, N. J., and C. Eroglu. 2017. 'Cell Biology of Astrocyte-Synapse Interactions', *Neuron*, 96: 697-708.

Angulo, M. C., A. S. Kozlov, S. Charpak, and E. Audinat. 2004. 'Glutamate released from glial cells synchronizes neuronal activity in the hippocampus', *J Neurosci*, 24: 6920-7.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. 'Gene ontology: tool for the unification of biology. The Gene Ontology Consortium', *Nat Genet*, 25: 25-9.

Awasaki, T., S. L. Lai, K. Ito, and T. Lee. 2008. 'Organization and postembryonic development of glial cells in the adult central brain of Drosophila', *J Neurosci*, 28: 13742-53.

Bailey, A. P., G. Koster, C. Guillermier, E. M. Hirst, J. I. MacRae, C. P. Lechene, A. D. Postle, and A. P. Gould. 2015. 'Antioxidant Role for Lipid Droplets in a Stem Cell Niche of Drosophila', *Cell*, 163: 340-53.

Bainton, R. J., L. T. Tsai, T. Schwabe, M. DeSalvo, U. Gaul, and U. Heberlein. 2005. 'moody encodes two GPCRs that regulate cocaine behaviors and blood-brain barrier permeability in Drosophila', *Cell*, 123: 145-56.

Barres, B. A. 2008. 'The mystery and magic of glia: a perspective on their roles in health and disease', *Neuron*, 60: 430-40.

Bayraktar, O. A., T. Bartels, S. Holmqvist, V. Kleshchevnikov, A. Martirosyan, D. Polioudakis, L. Ben Haim, A. M. H. Young, M. Y. Batiuk, K. Prakash, A. Brown, K. Roberts, M. F. Paredes, R. Kawaguchi, J. H. Stockley, K. Sabeur, S. M. Chang, E. Huang, P. Hutchinson, E. M. Ullian, M. Hemberg, G. Coppola, M. G. Holt, D. H. Geschwind, and D. H. Rowitch. 2020. 'Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map', *Nat Neurosci*, 23: 500-09.

Beckervordersandforth, R. M., C. Rickert, B. Altenhein, and G. M. Technau. 2008. 'Subtypes of glial cells in the Drosophila embryonic ventral nerve cord as related to lineage and gene expression', *Mech Dev*, 125: 542-57.

Benediktsson, A. M., G. S. Marrs, J. C. Tu, P. F. Worley, J. D. Rothstein, D. E. Bergles, and M. E. Dailey. 2012. 'Neuronal activity regulates glutamate transporter dynamics in developing astrocytes', *Glia*, 60: 175-88.

Boulanger, L. M., and C. J. Shatz. 2004. 'Immune signalling in neural development, synaptic plasticity and disease', *Nat Rev Neurosci*, 5: 521-31.

Brown, A. M., and B. R. Ransom. 2007. 'Astrocyte glycogen and brain energy metabolism', *Glia*, 55: 1263-71.

Bushong, E. A., M. E. Martone, and M. H. Ellisman. 2004. 'Maturation of astrocyte morphology and the establishment of astrocyte domains during postnatal hippocampal development', *Int J Dev Neurosci*, 22: 73-86.

Campos, A. R., D. R. Rosen, S. N. Robinow, and K. White. 1987. 'Molecular analysis of the locus elav in Drosophila melanogaster: a gene whose embryonic expression is neural specific', *EMBO J*, 6: 425-31.

Carlson, S. D., J. L. Juang, S. L. Hilgers, and M. B. Garment. 2000. 'Blood barriers of the insect', *Annu Rev Entomol*, 45: 151-74.

Chaboub, L. S., and B. Deneen. 2012. 'Developmental origins of astrocyte heterogeneity: the final frontier of CNS development', *Dev Neurosci*, 34: 379-88.

Clasadonte, J., J. Dong, D. J. Hines, and P. G. Haydon. 2013. 'Astrocyte control of synaptic NMDA receptors contributes to the progressive development of temporal lobe epilepsy', *Proc Natl Acad Sci U S A*, 110: 17540-5.

Cocanougher, Benjamin T., Jason D. Wittenbach, Xi Salina Long, Andrea B. Kohn, Tigran P. Norekian, Jinyao Yan, Jennifer Colonell, Jean-Baptiste Masson, James W. Truman, Albert Cardona, Srinivas C. Turaga, Robert H. Singer, Leonid L. Moroz, and Marta Zlatic. 2020. 'Comparative single-cell transcriptomics of complete insect nervous systems', *bioRxiv*: 785931.

Colombani, J., S. Raisin, S. Pantalacci, T. Radimerski, J. Montagne, and P. Leopold. 2003. 'A nutrient sensor mechanism controls Drosophila growth', *Cell*, 114: 739-49.

Colon-Ramos, D. A., M. A. Margeta, and K. Shen. 2007. 'Glia promote local synaptogenesis through UNC-6 (netrin) signaling in C. elegans', *Science*, 318: 103-6.

Cremazy, F., P. Berta, and F. Girard. 2000. 'Sox neuro, a new Drosophila Sox gene expressed in the developing central nervous system', *Mech Dev*, 93: 215-9.

Croset, V., C. D. Treiber, and S. Waddell. 2018. 'Cellular diversity in the Drosophila midbrain revealed by single-cell transcriptomics', *Elife*, 7.

Davie, K., J. Janssens, D. Koldere, M. De Waegeneer, U. Pech, L. Kreft, S. Aibar, S. Makhzami, V. Christiaens, C. Bravo Gonzalez-Blas, S. Poovathingal, G. Hulselmans, K. I. Spanier, T. Moerman, B. Vanspauwen, S. Geurs, T. Voet, J. Lammertyn, B. Thienpont, S. Liu, N. Konstantinides, M. Fiers, P. Verstreken, and S. Aerts. 2018. 'A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain', *Cell*, 174: 982-98 e20.

DeSalvo, M. K., S. J. Hindle, Z. M. Rusan, S. Orng, M. Eddison, K. Halliwill, and R. J. Bainton. 2014. 'The Drosophila surface glia transcriptome: evolutionary conserved blood-brain barrier processes', *Front Neurosci*, 8: 346.

Devaraju, P., M. Y. Sun, T. L. Myers, K. Lauderdale, and T. A. Fiacco. 2013. 'Astrocytic group I mGluR-dependent potentiation of astrocytic glutamate and potassium uptake', *J Neurophysiol*, 109: 2404-14.

DiAntonio, A., R. W. Burgess, A. C. Chin, D. L. Deitcher, R. H. Scheller, and T. L. Schwarz. 1993. 'Identification and characterization of Drosophila genes for synaptic vesicle proteins', *J Neurosci*, 13: 4924-35.

Diao, F., H. Ironfield, H. Luan, F. Diao, W. C. Shropshire, J. Ewer, E. Marr, C. J. Potter, M. Landgraf, and B. H. White. 2015. 'Plug-and-play genetic access to drosophila cell types using exchangeable exon cassettes', *Cell Rep*, 10: 1410-21.

Distler, C., Z. Dreher, and J. Stone. 1991. 'Contact spacing among astrocytes in the central nervous system: an hypothesis of their structural role', *Glia*, 4: 484-94.

Doherty, J., M. A. Logan, O. E. Tasdemir, and M. R. Freeman. 2009. 'Ensheathing glia function as phagocytes in the adult Drosophila brain', *J Neurosci*, 29: 4768-81.

Dumstrei, K., F. Wang, and V. Hartenstein. 2003. 'Role of DE-cadherin in neuroblast proliferation, neural morphogenesis, and axon tract formation in Drosophila larval brain development', *J Neurosci*, 23: 3325-35.

Duran, J., M. F. Tevy, M. Garcia-Rocha, J. Calbo, M. Milan, and J. J. Guinovart. 2012. 'Deleterious effects of neuronal accumulation of glycogen in flies and mice', *EMBO Mol Med*, 4: 719-29.

Ebens, A. J., H. Garren, B. N. Cheyette, and S. L. Zipursky. 1993. 'The Drosophila anachronism locus: a glycoprotein secreted by glia inhibits neuroblast proliferation', *Cell*, 74: 15-27.

Edgar, B. A., F. Sprenger, R. J. Duronio, P. Leopold, and P. H. O'Farrell. 1994. 'Distinct molecular mechanism regulate cell cycle timing at successive stages of Drosophila embryogenesis', *Genes Dev*, 8: 440-52.

Eichler, K., F. Li, A. Litwin-Kumar, Y. Park, I. Andrade, C. M. Schneider-Mizell, T. Saumweber, A. Huser, C. Eschbach, B. Gerber, R. D. Fetter, J. W. Truman, C. E. Priebe, L. F. Abbott, A. S. Thum, M. Zlatic, and A. Cardona. 2017. 'The complete connectome of a learning and memory centre in an insect brain', *Nature*, 548: 175-82.

Enriquez, J., L. Q. Rio, R. Blazeski, S. Bellemin, P. Godement, C. Mason, and R. S. Mann. 2018. 'Differing Strategies Despite Shared Lineages of Motor Neurons and Glia to Achieve Robust Development of an Adult Neuropil in Drosophila', *Neuron*, 97: 538-54 e5.

Featherstone, D. E. 2011. 'Glial solute carrier transporters in Drosophila and mice', *Glia*, 59: 1351-63.

Fields, R. D. 2006. 'Advances in understanding neuron-glia interactions', *Neuron Glia Biol*, 2: 23-6.

Freeman, M. R. 2015. 'Drosophila Central Nervous System Glia', *Cold Spring Harb Perspect Biol*, 7.

Freeman, M. R., J. Delrow, J. Kim, E. Johnson, and C. Q. Doe. 2003. 'Unwrapping glial biology: Gcm target genes regulating glial development, diversification, and function', *Neuron*, 38: 567-80.

Freeman, M. R., and J. Doherty. 2006. 'Glial cell biology in Drosophila and vertebrates', *Trends Neurosci*, 29: 82-90.

Gentry, M. S., J. J. Guinovart, B. A. Minassian, P. J. Roach, and J. M. Serratosa. 2018. 'Lafora disease offers a unique window into neuronal glycogen metabolism', *J Biol Chem*, 293: 7117-25.

Grimm, J. B., A. K. Muthusamy, Y. Liang, T. A. Brown, W. C. Lemon, R. Patel, R. Lu, J. J. Macklin, P. J. Keller, N. Ji, and L. D. Lavis. 2017. 'A general method to fine-tune fluorophores for live-cell and in vivo imaging', *Nat Methods*, 14: 987-94.

Halassa, M. M., C. Florian, T. Fellin, J. R. Munoz, S. Y. Lee, T. Abel, P. G. Haydon, and M. G. Frank. 2009. 'Astrocytic modulation of sleep homeostasis and cognitive consequences of sleep loss', *Neuron*, 61: 213-9.

Halter, D. A., J. Urban, C. Rickert, S. S. Ner, K. Ito, A. A. Travers, and G. M. Technau. 1995. 'The homeobox gene repo is required for the differentiation and maintenance of glia function in the embryonic nervous system of Drosophila melanogaster', *Development*, 121: 317-32.

Hartenstein, V. 2011. 'Morphological diversity and development of glia in Drosophila', *Glia*, 59: 1237-52.

Havula, E., M. Teesalu, T. Hyotylainen, H. Seppala, K. Hasygar, P. Auvinen, M. Oresic, T. Sandmann, and V. Hietakangas. 2013. 'Mondo/ChREBP-Mlx-regulated transcriptional network is essential for dietary sugar tolerance in Drosophila', *PLoS Genet*, 9: e1003438.

Hawkins, B. T., and T. P. Davis. 2005. 'The blood-brain barrier/neurovascular unit in health and disease', *Pharmacol Rev*, 57: 173-85.

Haydon, P. G., J. Blendy, S. J. Moss, and F. Rob Jackson. 2009. 'Astrocytic control of synaptic transmission and plasticity: a target for drugs of abuse?', *Neuropharmacology*, 56 Suppl 1: 83-90.

Haydon, P. G., and G. Carmignoto. 2006. 'Astrocyte control of synaptic transmission and neurovascular coupling', *Physiol Rev*, 86: 1009-31.

Hediger, M. A., M. F. Romero, J. B. Peng, A. Rolfs, H. Takanaga, and E. A. Bruford. 2004. 'The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction', *Pflugers Arch*, 447: 465-8.

Helmstaedter, M., K. L. Briggman, S. C. Turaga, V. Jain, H. S. Seung, and W. Denk. 2013. 'Connectomic reconstruction of the inner plexiform layer in the mouse retina', *Nature*, 500: 168-74.

Hildebrand, D. G. C., M. Cicconet, R. M. Torres, W. Choi, T. M. Quan, J. Moon, A. W. Wetzel, A. Scott Champion, B. J. Graham, O. Randlett, G. S. Plummer, R. Portugues, I. H. Bianco, S. Saalfeld, A. D. Baden, K. Lillaney, R. Burns, J. T. Vogelstein, A. F. Schier, W. A. Lee, W. K. Jeong, J. W. Lichtman, and F. Engert. 2017. 'Whole-brain serial-section electron microscopy in larval zebrafish', *Nature*, 545: 345-49.

Hosoya, T., K. Takizawa, K. Nitta, and Y. Hotta. 1995. 'glial cells missing: a binary switch between neuronal and glial determination in Drosophila', *Cell*, 82: 1025-36.

Hoyle, G., M. Williams, and C. Phillips. 1986. 'Functional morphology of insect neuronal cell-surface/glial contacts: the trophospongium', *J Comp Neurol*, 246: 113-28.

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

Huang, Y., F. S. Ng, and F. R. Jackson. 2015. 'Comparison of larval and adult Drosophila astrocytes reveals stage-specific gene expression profiles', *G3 (Bethesda)*, 5: 551-8.

Izumi, Y., M. Motoishi, K. Furuse, and M. Furuse. 2016. 'A tetraspanin regulates septate junction formation in Drosophila midgut', *J Cell Sci*, 129: 1155-64.

Jackson, F. R. 2011. 'Glial cell modulation of circadian rhythms', *Glia*, 59: 1341-50.

Jarrell, T. A., Y. Wang, A. E. Bloniarz, C. A. Brittin, M. Xu, J. N. Thomson, D. G. Albertson, D. H. Hall, and S. W. Emmons. 2012. 'The connectome of a decision-making neural network', *Science*, 337: 437-44.

Jones, B. W., R. D. Fetter, G. Tear, and C. S. Goodman. 1995. 'glial cells missing: a genetic switch that controls glial versus neuronal fate', *Cell*, 82: 1013-23.

Kato, K., M. Orihara-Ono, and T. Awasaki. 2020. 'Multiple lineages enable robust development of the neuropil-glia architecture in adult Drosophila', *Development*, 147.

Kessaris, N., N. Pringle, and W. D. Richardson. 2008. 'Specification of CNS glia from neural stem cells in the embryonic neuroepithelium', *Philos Trans R Soc Lond B Biol Sci*, 363: 71-85.

Kis, V., B. Barti, M. Lippai, and M. Sass. 2015. 'Specialized Cortex Glial Cells Accumulate Lipid Droplets in Drosophila melanogaster', *PLoS One*, 10: e0131250.

Konstantinides, N., K. Kapuralin, C. Fadil, L. Barboza, R. Satija, and C. Desplan. 2018. 'Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features', *Cell*, 174: 622-35 e13.

Kremer, M. C., C. Jung, S. Batelli, G. M. Rubin, and U. Gaul. 2017. 'The glia of the adult Drosophila nervous system', *Glia*, 65: 606-38.

Lefevre, P., A. Rochat, C. Bodemer, P. Vabres, Y. Barrandon, Y. de Prost, C. Garner, and A. Hovnanian. 2000. 'Linkage of Marie-Unna hypotrichosis locus to chromosome 8p21 and exclusion of 10 genes including the hairless gene by mutation analysis', *Eur J Hum Genet*, 8: 273-9.

Lemon, W. C., S. R. Pulver, B. Hockendorf, K. McDole, K. Branson, J. Freeman, and P. J. Keller. 2015. 'Whole-central nervous system functional imaging in larval Drosophila', *Nat Commun*, 6: 7924.

Li, C., J. Zheng, S. Chen, B. Huang, G. Li, Z. Feng, J. Wang, and S. Xu. 2018. 'RRM2 promotes the progression of human glioblastoma', *J Cell Physiol*, 233: 6759-67.

Li, H. H., J. R. Kroll, S. M. Lennox, O. Ogundeyi, J. Jeter, G. Depasquale, and J. W. Truman. 2014. 'A GAL4 driver resource for developmental and behavioral studies on the larval CNS of Drosophila', *Cell Rep*, 8: 897-908.

Lissemore, J. L., C. A. Baumgardner, B. W. Geer, and D. T. Sullivan. 1990. 'Effect of dietary carbohydrates and ethanol on expression of genes encoding sn-glycerol-3-phosphate dehydrogenase, aldolase, and phosphoglycerate kinase in Drosophila larvae', *Biochem Genet*, 28: 615-30.

Liu, L., K. R. MacKenzie, N. Putluri, M. Maletic-Savatic, and H. J. Bellen. 2017. 'The Glia-Neuron Lactate Shuttle and Elevated ROS Promote Lipid Synthesis in Neurons and Lipid Droplet Accumulation in Glia via APOE/D', *Cell Metab*, 26: 719-37 e6.

Liu, L., K. Zhang, H. Sandoval, S. Yamamoto, M. Jaiswal, E. Sanz, Z. Li, J. Hui, B. H. Graham, A. Quintana, and H. J. Bellen. 2015. 'Glial lipid droplets and ROS induced by mitochondrial defects promote neurodegeneration', *Cell*, 160: 177-90.

MacDonald, J. M., M. G. Beach, E. Porpiglia, A. E. Sheehan, R. J. Watts, and M. R. Freeman. 2006. 'The Drosophila cell corpse engulfment receptor Draper mediates glial clearance of severed axons', *Neuron*, 50: 869-81.

MacNamee, S. E., K. E. Liu, S. Gerhard, C. T. Tran, R. D. Fetter, A. Cardona, L. P. Tolbert, and L. A. Oland. 2016. 'Astrocytic glutamate transport regulates a Drosophila CNS synapse that lacks astrocyte ensheathment', *J Comp Neurol*, 524: 1979-98.

Maecker, H. T., S. C. Todd, and S. Levy. 1997. 'The tetraspanin superfamily: molecular facilitators', *FASEB J*, 11: 428-42.

Melom, J. E., and J. T. Littleton. 2013. 'Mutation of a NCKX eliminates glial microdomain calcium oscillations and enhances seizure susceptibility', *J Neurosci*, 33: 1169-78.

Molofsky, A. V., K. W. Kelley, H. H. Tsai, S. A. Redmond, S. M. Chang, L. Madireddy, J. R. Chan, S. E. Baranzini, E. M. Ullian, and D. H. Rowitch. 2014. 'Astrocyte-encoded positional cues maintain sensorimotor circuit integrity', *Nature*, 509: 189-94.

Muthukumar, A. K., T. Stork, and M. R. Freeman. 2014. 'Activity-dependent regulation of astrocyte GAT levels during synaptogenesis', *Nat Neurosci*, 17: 1340-50.

Ng, F. S., S. Sengupta, Y. Huang, A. M. Yu, S. You, M. A. Roberts, L. K. Iyer, Y. Yang, and F. R. Jackson. 2016. 'TRAP-seq Profiling and RNAi-Based Genetic Screens Identify Conserved Glial Genes Required for Adult Drosophila Behavior', *Front Mol Neurosci*, 9: 146.

Nishiyama, A., R. Suzuki, and X. Zhu. 2014. 'NG2 cells (polydendrocytes) in brain physiology and repair', *Front Neurosci*, 8: 133.

Noordermeer, J. N., C. C. Kopczynski, R. D. Fetter, K. S. Bland, W. Y. Chen, and C. S. Goodman. 1998. 'Wrapper, a novel member of the Ig superfamily, is expressed by midline glia and is required for them to ensheath commissural axons in Drosophila', *Neuron*, 21: 991-1001.

Ogata, K., and T. Kosaka. 2002. 'Structural and quantitative analysis of astrocytes in the mouse hippocampus', *Neuroscience*, 113: 221-33.

Ohyama, T., C. M. Schneider-Mizell, R. D. Fetter, J. V. Aleman, R. Franconville, M. Rivera-Alba, B. D. Mensh, K. M. Branson, J. H. Simpson, J. W. Truman, A. Cardona, and M. Zlatic. 2015. 'A multilevel multimodal circuit enhances action selection in Drosophila', *Nature*, 520: 633-9.

Omoto, J. J., J. K. Lovick, and V. Hartenstein. 2016. 'Origins of glial cell populations in the insect nervous system', *Curr Opin Insect Sci*, 18: 96-104.

Panatier, A., D. T. Theodosis, J. P. Mothet, B. Touquet, L. Pollegioni, D. A. Poulain, and S. H. Oliet. 2006. 'Glia-derived D-serine controls NMDA receptor activity and synaptic memory', *Cell*, 125: 775-84.

Parkhurst, S. J., P. Adhikari, J. S. Navarrete, A. Legendre, M. Manansala, and F. W. Wolf. 2018. 'Perineurial Barrier Glia Physically Respond to Alcohol in an Akap200-Dependent Manner to Promote Tolerance', *Cell Rep*, 22: 1647-56.

Pascual, O., K. B. Casper, C. Kubera, J. Zhang, R. Revilla-Sanchez, J. Y. Sul, H. Takano, S. J. Moss, K. McCarthy, and P. G. Haydon. 2005. 'Astrocytic purinergic signaling coordinates synaptic networks', *Science*, 310: 113-6.

Peco, E., S. Davla, D. Camp, S. M. Stacey, M. Landgraf, and D. J. van Meyel. 2016. 'Drosophila astrocytes cover specific territories of the CNS neuropil and are instructed to differentiate by Prospero, a key effector of Notch', *Development*, 143: 1170-81.

Pereanu, W., A. Kumar, A. Jennett, H. Reichert, and V. Hartenstein. 2010. 'Development-based compartmentalization of the Drosophila central brain', *J Comp Neurol*, 518: 2996-3023.

Pereanu, W., D. Shy, and V. Hartenstein. 2005. 'Morphogenesis and proliferation of the larval brain glia in Drosophila', *Dev Biol*, 283: 191-203.

Pereanu, W., S. Spindler, L. Cruz, and V. Hartenstein. 2007. 'Tracheal development in the Drosophila brain is constrained by glial cells', *Dev Biol*, 302: 169-80.

Piggott, B. J., C. J. Peters, Y. He, X. Huang, S. Younger, L. Y. Jan, and Y. N. Jan. 2019. 'Paralytic, the Drosophila voltage-gated sodium channel, regulates proliferation of neural progenitors', *Genes Dev*, 33: 1739-50.

Plazaola-Sasieta, H., Q. Zhu, H. Gaitan-Penas, M. Rios, R. Estevez, and M. Morey. 2019. 'Drosophila ClC-a is required in glia of the stem cell niche for proper neurogenesis and wiring of neural circuits', *Glia*, 67: 2374-98.

Pocock, J. M., and H. Kettenmann. 2007. 'Neurotransmitter receptors on microglia', *Trends Neurosci*, 30: 527-35.

Prevedel, R., Y. G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrodel, R. Raskar, M. Zimmer, E. S. Boyden, and A. Vaziri. 2014. 'Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy', *Nat Methods*, 11: 727-30.

Ren, Q., K. Chen, and I. T. Paulsen. 2007. 'TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels', *Nucleic Acids Res*, 35: D274-9.

Robinow, S., and K. White. 1988. 'The locus elav of Drosophila melanogaster is expressed in neurons at all developmental stages', *Dev Biol*, 126: 294-303.

Rusnakova, V., P. Honsa, D. Dzamba, A. Stahlberg, M. Kubista, and M. Anderova. 2013. 'Heterogeneity of astrocytes: from development to injury - single cell gene expression', *PLoS One*, 8: e69734.

Salter, M. W., and B. Stevens. 2017. 'Microglia emerge as central players in brain disease', *Nat Med*, 23: 1018-27.

Satija, R., J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev. 2015. 'Spatial reconstruction of single-cell gene expression data', *Nat Biotechnol*, 33: 495-502.

Saunders, A., E. Z. Macosko, A. Wysoker, M. Goldman, F. M. Krienen, H. de Rivera, E. Bien, M. Baum, L. Bortolin, S. Wang, A. Goeva, J. Nemesh, N. Kamitaki, S. Brumbaugh, D. Kulp, and S. A. McCarroll. 2018. 'Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain', *Cell*, 174: 1015-30 e16.

Schneider-Mizell, C. M., S. Gerhard, M. Longair, T. Kazimiers, F. Li, M. F. Zwart, A. Champion, F. M. Midgley, R. D. Fetter, S. Saalfeld, and A. Cardona. 2016. 'Quantitative neuroanatomy for connectomics in Drosophila', *Elife*, 5.

Schwabe, T., R. J. Bainton, R. D. Fetter, U. Heberlein, and U. Gaul. 2005. 'GPCR signaling is required for blood-brain barrier formation in drosophila', *Cell*, 123: 133-44.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. 'Cytoscape: a software environment for integrated models of biomolecular interaction networks', *Genome Res*, 13: 2498-504.

Shigetomi, E., D. N. Bowser, M. V. Sofroniew, and B. S. Khakh. 2008. 'Two forms of astrocyte calcium excitability have distinct effects on NMDA receptor-mediated slow inward currents in pyramidal neurons', *J Neurosci*, 28: 6659-63.

Sofroniew, M. V., and H. V. Vinters. 2010. 'Astrocytes: biology and pathology', *Acta Neuropathol*, 119: 7-35.

Sonnenfeld, M. J., and J. R. Jacobs. 1995. 'Macrophages and glia participate in the removal of apoptotic neurons from the Drosophila embryonic nervous system', *J Comp Neurol*, 359: 644-52.

Spindler, S. R., I. Ortiz, S. Fung, S. Takashima, and V. Hartenstein. 2009. 'Drosophila cortex and neuropile glia influence secondary axon tract growth, pathfinding, and fasciculation in the developing larval brain', *Dev Biol*, 334: 355-68.

Stacey, S. M., N. I. Muraro, E. Peco, A. Labbe, G. B. Thomas, R. A. Baines, and D. J. van Meyel. 2010. 'Drosophila glial glutamate transporter Eaat1 is regulated by fringe-mediated notch signaling and is essential for larval locomotion', *J Neurosci*, 30: 14446-57.

Stacey, S. M., G. B. Thomas, A. Labbe, and D. J. Van Meyel. 2007. 'Longitudinal glia in the fly CNS: pushing the envelope on glial diversity and neuron-glial interactions', *Neuron Glia Biol*, 3: 27-33.

Stacey, Staphanie. 2010. 'Drosophila Glia Glutamate Transporter Eaat1 ia required by Fringe-mediated Notch signalling and is essential for Larval Locomotion', *Journal of Neuroscience*.

Stahlberg, A., D. Andersson, J. Aurelius, M. Faiz, M. Pekna, M. Kubista, and M. Pekny. 2011. 'Defining cell populations with single-cell gene expression profiling: correlations and identification of astrocyte subpopulations', *Nucleic Acids Res*, 39: e24.

Stork, T., R. Bernardos, and M. R. Freeman. 2012. 'Analysis of glial cell development and function in Drosophila', *Cold Spring Harb Protoc*, 2012: 1-17.

Stork, T., D. Engelen, A. Krudewig, M. Silies, R. J. Bainton, and C. Klambt. 2008. 'Organization and function of the blood-brain barrier in Drosophila', *J Neurosci*, 28: 587-97.

Stork, T., A. Sheehan, O. E. Tasdemir-Yilmaz, and M. R. Freeman. 2014. 'Neuron-glia interactions through the Heartless FGF receptor signaling pathway mediate morphogenesis of Drosophila astrocytes', *Neuron*, 83: 388-403.

Sun, B., and P. M. Salvaterra. 1995. 'Two Drosophila nervous system antigens, Nervana 1 and 2, are homologous to the beta subunit of Na+,K(+)-ATPase', *Proc Natl Acad Sci U S A*, 92: 5396-400.

Takemura, S. Y., A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, T. Zhao, J. A. Horne, R. D. Fetter, S. Takemura, K. Blazek, L. A. Chang, O. Ogundeyi, M. A. Saunders, V. Shapiro, C. Sigmund, G. M. Rubin, L. K. Scheffer, I. A. Meinertzhagen, and D. B. Chklovskii. 2013. 'A visual motion detection circuit suggested by Drosophila connectomics', *Nature*, 500: 175-81.

Tepass, U., and V. Hartenstein. 1994. 'The development of cellular junctions in the Drosophila embryo', *Dev Biol*, 161: 563-96.

Tietz, S., and B. Engelhardt. 2015. 'Brain barriers: Crosstalk between complex tight junctions and adherens junctions', *J Cell Biol*, 209: 493-506.

Trotter, J., K. Karram, and A. Nishiyama. 2010. 'NG2 cells: Properties, progeny and origin', *Brain Res Rev*, 63: 72-82.

Venken, Koen JT. 2011. 'MiMIC: a highly versatile transposoon insertion resource for engineering Drosophila melanogaster genes', *Nature Methods*.

Vincent, S., J. L. Vonesch, and A. Giangrande. 1996. 'Glide directs glial fate commitment and cell fate switch between neurones and glia', *Development*, 122: 131-9.

Wagh, D. A., T. M. Rasse, E. Asan, A. Hofbauer, I. Schwenkert, H. Durrbeck, S. Buchner, M. C. Dabauvalle, M. Schmidt, G. Qin, C. Wichmann, R. Kittel, S. J. Sigrist, and E. Buchner. 2006. 'Bruchpilot, a protein with homology to ELKS/CAST, is required for structural integrity and function of synaptic active zones in Drosophila', *Neuron*, 49: 833-44.

Xie, M., D. Kamenev, M. Kaucka, M. E. Kastriti, B. Zhou, A. V. Artemov, M. Storer, K. Fried, I. Adameyko, V. Dyachuk, and A. S. Chagin. 2019. 'Schwann cell precursors contribute to skeletal formation during embryonic development in mice and zebrafish', *Proc Natl Acad Sci U S A*, 116: 15068-73.

Xiong, W. C., H. Okano, N. H. Patel, J. A. Blendy, and C. Montell. 1994. 'repo encodes a glial-specific homeo domain protein required in the Drosophila nervous system', *Genes Dev*, 8: 981-94.

Xu, J., A. M. Bernstein, A. Wong, X. H. Lu, S. Khoja, X. W. Yang, D. L. Davies, P. Micevych, M. V. Sofroniew, and B. S. Khakh. 2016. 'P2X4 Receptor Reporter Mice: Sparse Brain Expression and Feeding-Related Presynaptic Facilitation in the Arcuate Nucleus', *J Neurosci*, 36: 8902-20.

Yang, Y., O. Gozen, A. Watkins, I. Lorenzini, A. Lepore, Y. Gao, S. Vidensky, J. Brennan, D. Poulsen, J. Won Park, N. Li Jeon, M. B. Robinson, and J. D. Rothstein. 2009. 'Presynaptic regulation of astroglial excitatory neurotransmitter transporter GLT1', *Neuron*, 61: 880-94.

Yildirim, K., J. Petri, R. Kottmeier, and C. Klambt. 2019. 'Drosophila glia: Few cell types and many conserved functions', *Glia*, 67: 5-26.

Zelzer, E., P. Wappner, and B. Shilo. 1997. 'The PAS domain confers target gene specificity of *Drosophila* bHLH-PAS proteins', *Genes Dev.*, 11: 2079-89.

Zhang, Y., and B. A. Barres. 2010. 'Astrocyte heterogeneity: an underappreciated topic in neurobiology', *Curr Opin Neurobiol*, 20: 588-94.

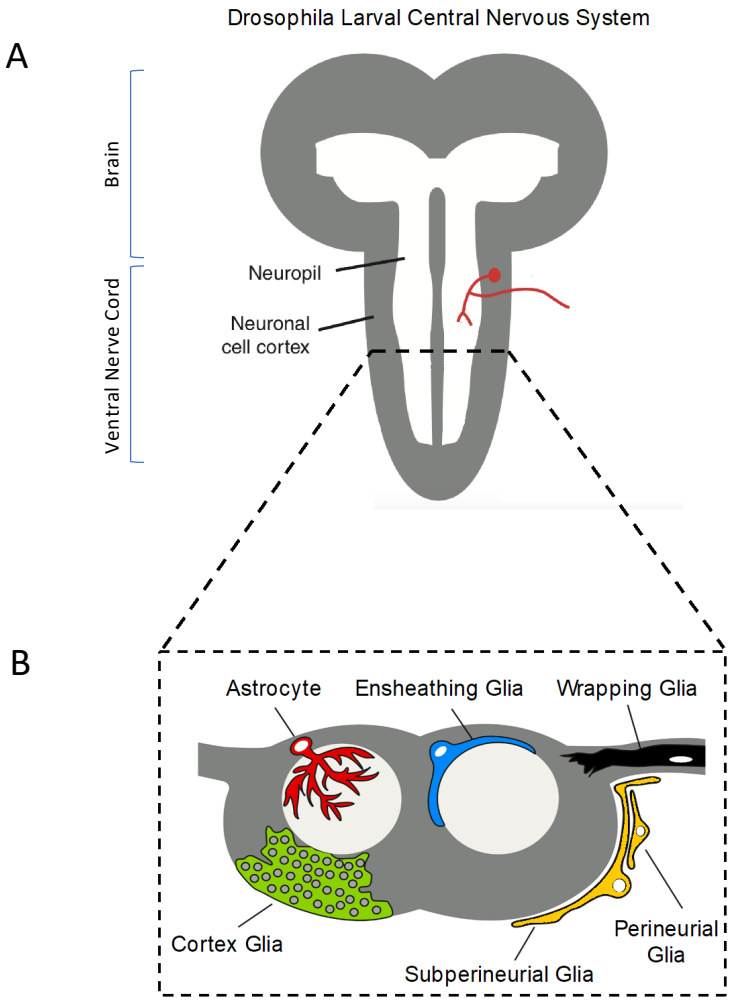
Zhang, Y. V., K. G. Ormerod, and J. T. Littleton. 2017. 'Astrocyte Ca(2+) Influx Negatively Regulates Neuronal Activity', *eNeuro*, 4.

Zheng, G. X., J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M.

Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, and J. H. Bielas. 2017. 'Massively parallel digital transcriptional profiling of single cells', *Nat Commun*, 8: 14049.

Zhou, X., W. J. Liao, J. M. Liao, P. Liao, and H. Lu. 2015. 'Ribosomal proteins: functions beyond the ribosome', *J Mol Cell Biol*, 7: 92-104.

Figures


A
Drosophila Larval Central Nervous System
Brain
Ventral Nerve Cord
Neuropil
Neuronal cell cortex
B
Astrocyte
Ensheathing Glia
Wrapping Glia
Cortex Glia
Perineurial Glia
Subperineurial Glia

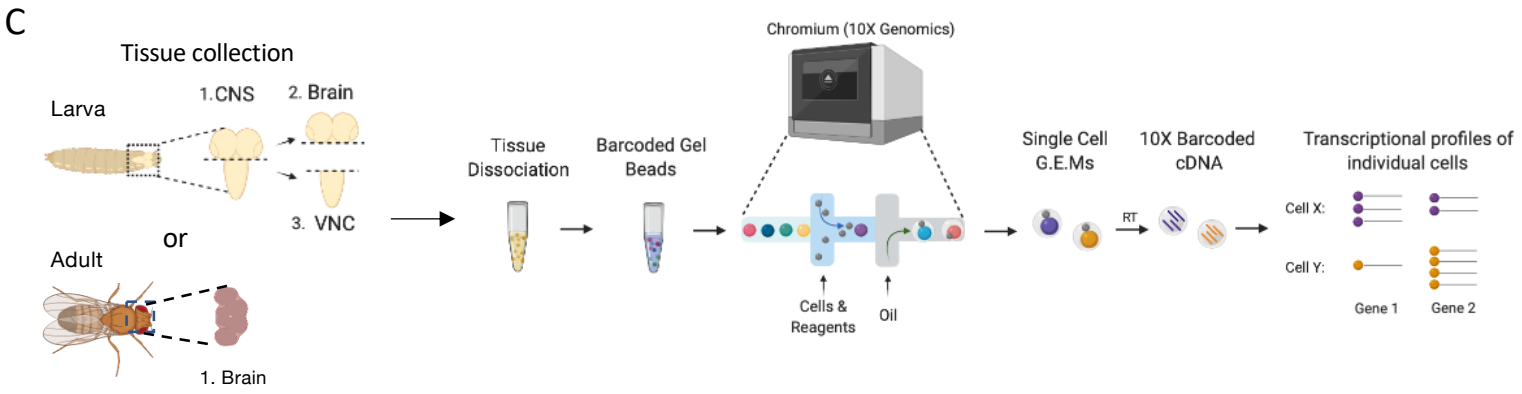
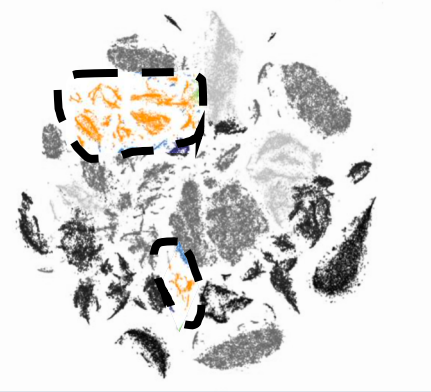

C
Tissue collection
Larva
1.CNS
2. Brain
3. VNC
or
Adult
1. Brain
Tissue Dissociation
Barcoded Gel Beads
Chromium (10X Genomics)
Cells & Reagents
Oil
Single Cell G.E.Ms
10X Barcoded cDNA
RT
Transcriptional profiles of individual cells
Cell X:
Cell Y:
Gene 1
Gene 2

Figure 1. Glia within the Drosophila larval CNS captured via scRNA sequencing
(A) Schematic diagram of larval CNS, showing the neuronal cell cortex (grey), neuropil (white), in the brain lobes (top) and ventral nerve cord (VNC) (bottom). Neuron cell bodies lie in the neuronal cell cortex and their synapses are formed in the neuropil (red). (B) Cross-sectional schematic view of the CNS to describe distinct glial subtypes: astrocytes (red), ensheathing glia (blue), cortex glia (green), surface glia (yellow). Wrapping glia (black) are found in the nerves carrying motor and sensory axons to and from the periphery and are often not included in CNS dissections. Morphological arrangement in the adult brain is similar. Adapted from (Freeman 2015). (C) Workflow of dissections and the 10X Chromium scRNAseq process for both larval and adult CNS tissues to gather matrix data sets of gene expression.

A

t-SNE plot of clusters in the Larval 100K data set



B

Glia Extraction

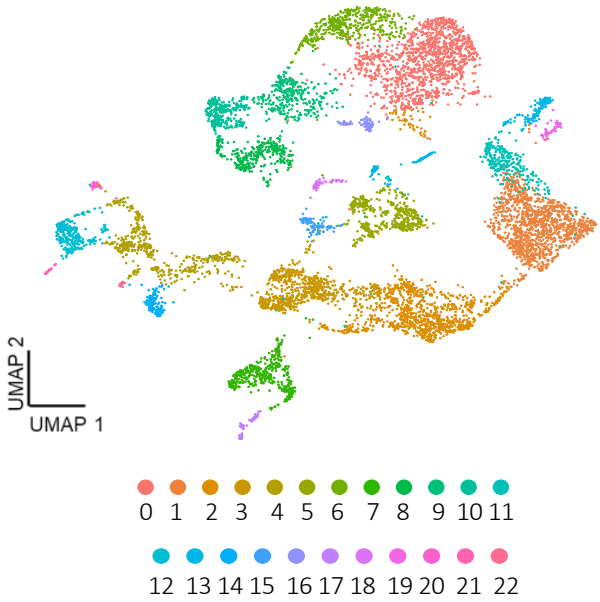
12,068 genes
x
8,566 cells

Quality Control & pre-processing

9,625 genes
x
8,531 cells

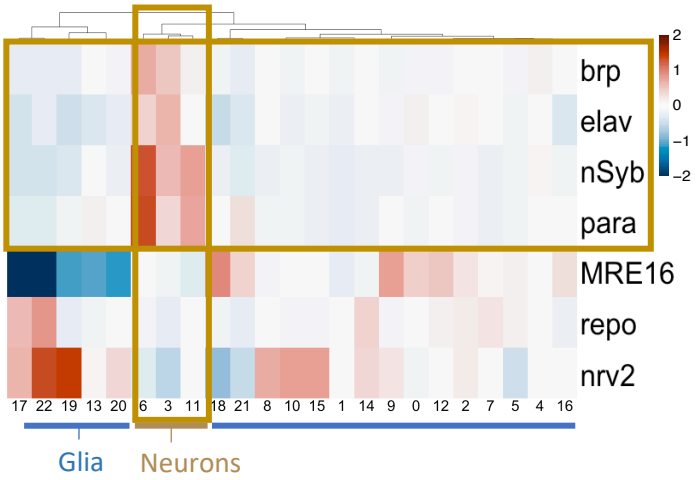
C

Larval 8.5K UMAP



D

Larval 8.5K Heatmap of neuronal and glial markers



23 Clusters (8.5K)

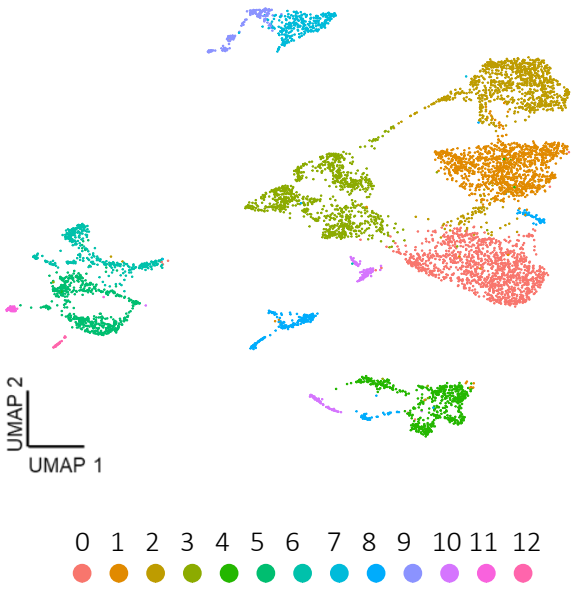
20 non-neuronal clusters

3 neuronal clusters

13 Clusters (7K)

E

Larval 7K UMAP



F

Larval 7K Heatmap of neuronal and glial markers

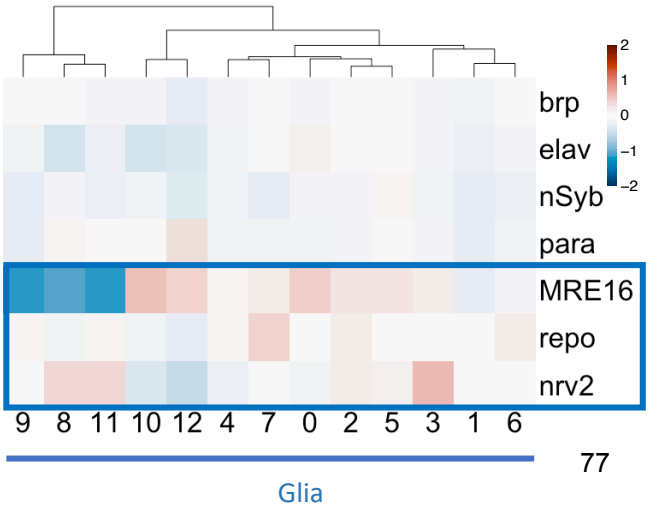
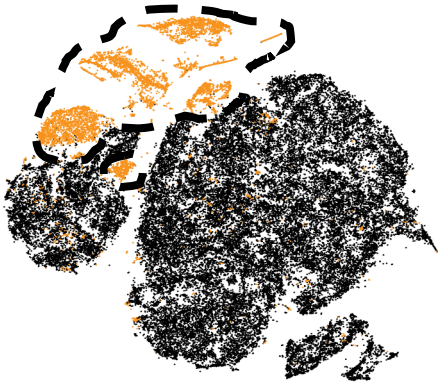


Figure 2: Glia-enrichment of larval data

(A) t-SNE plot from the Zlatic lab showing the entire starting larval CNS data set (100K) with glial cells highlighted (orange). (B) Extraction of glial cells followed by quality control and pre-processing reduced the number of cells and genes in the data set resulting in (C) UMAP of larval 8.5K data set with 23 cell populations using first 10 dimensions, 30 neighbors and resolution 0.6. (D) Heatmap of the mean scaled log-normalized expression, by cluster, of glial and neuronal marker genes, where red represents high expression and blue represents negative expression in the larval 8.5K data set. Clusters expressing an array of neuronal markers are indicated (box). These neuronal clusters were removed, resulting in (E) UMAP of final glia-enriched larval 7K data set with 13 cell populations using first 10 dimensions, 30 neighbors and resolution 0.2. (F) Heatmap of the mean scaled log-normalized expression, by cluster, of glial and neuronal marker genes where red represents high expression and blue represents negative expression in larval 7K data set. Glial markers are highlighted to show glia enrichment process was successful (box).

A

t-SNE plot of clusters in the Adult 57K data set



B

Glia Extraction

12,925 genes
x
7,910 cells

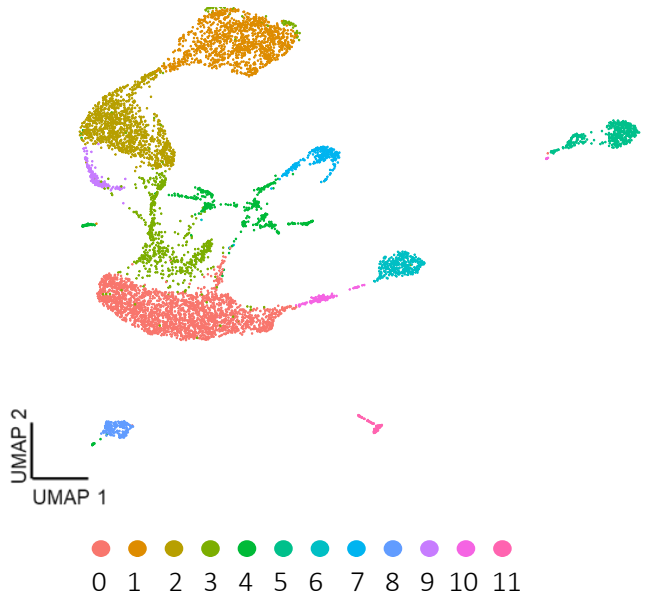
→

Quality Control & pre-processing

10,839 genes
x
7,642 cells

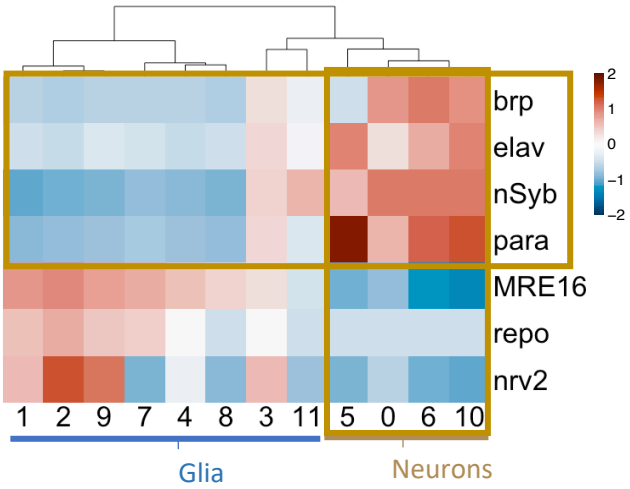
C

Adult 7.6K UMAP



D

Adult 7.6K Heatmap of neuronal and glial markers



12 Clusters (7.6K)

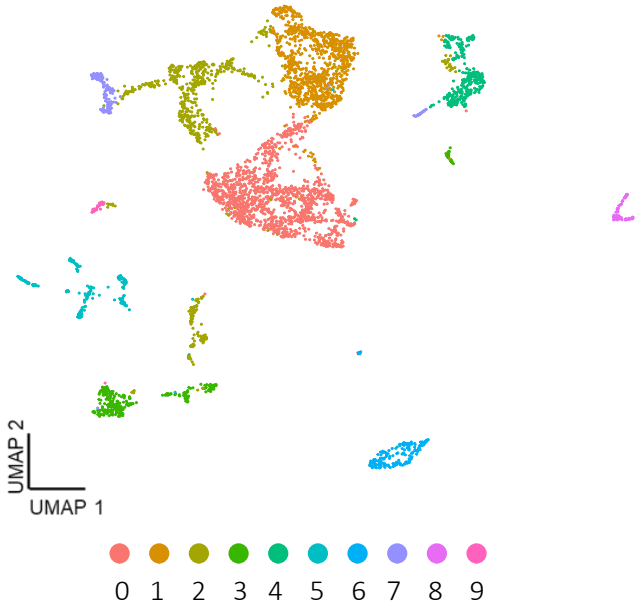
8 non-neuronal clusters

4 neuronal clusters

10 Clusters (4.4K)

E

Adult 4.4K UMAP



F

Adult 4.4K Heatmap of neuronal and glial markers

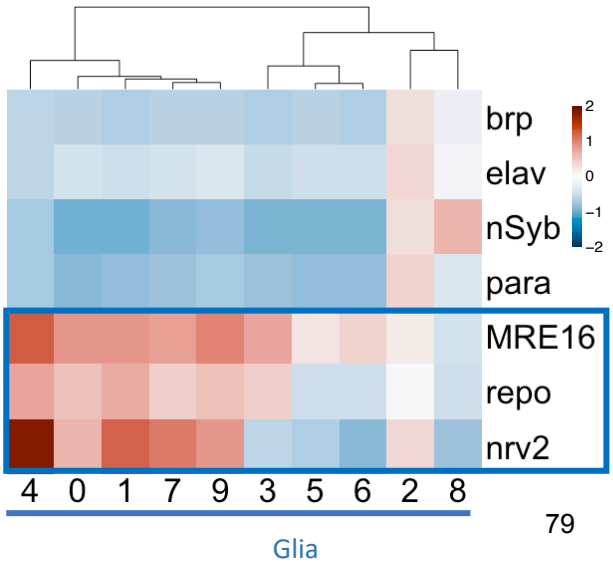


Figure 3: Glia-enrichment of adult data

(A) t-SNE plot from the Aerts lab showing the entire starting adult CNS data set (57K) with glial cells highlighted (orange) (B) Extraction of glial cells followed by quality control and pre-processing resulted in (C) UMAP of adult 7.6K data set with 12 cell populations using first 10 dimensions, 30 neighbors and resolution 0.2. (D) Heatmap of the mean scaled log-normalized expression, by cluster, of glial and neuronal marker genes where red represents high expression and blue represents negative expression. Clusters expressing an array of neuronal markers are indicated (box). These clusters were removed, resulting in (E) UMAP of final glia-enriched adult 4.4K data set with 10 cell populations using first 20 dimensions, 20 neighbors and resolution 0.1. (F) Heatmap of the mean scaled log-normalized expression, by cluster, of glial and neuronal marker genes where red represents high expression and blue represents negative expression, and glial markers are highlighted to show glia enrichment process was successful (box). Cluster 2 expresses both neuronal and glial markers, and cluster 8 expresses the neuronal marker nSyb.

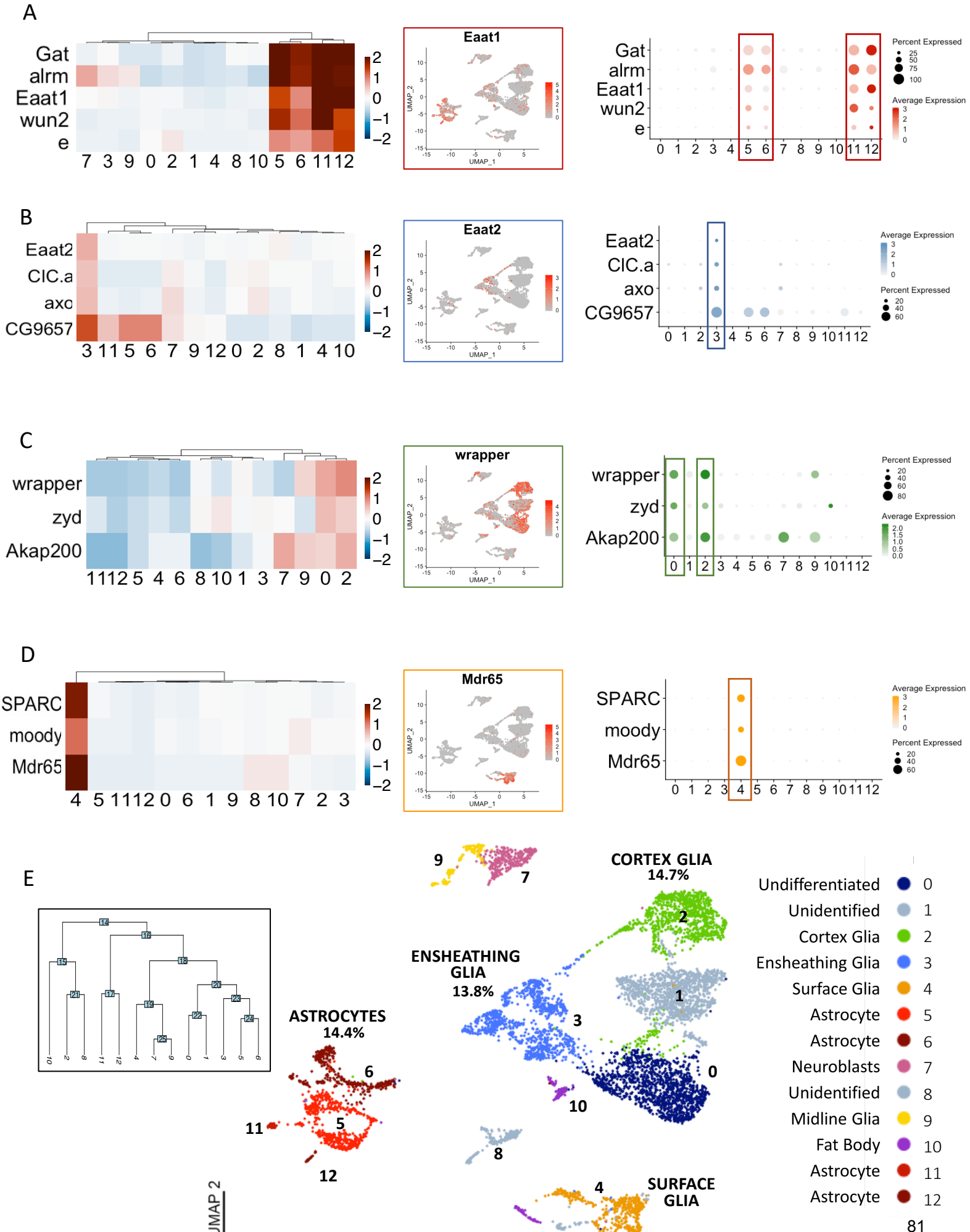

A
Gat
alrm
Eaat1
wun2
e
7 3 9 0 2 1 4 8 10 5 6 11 12
Eaat1
Percent Expressed
Average Expression
B
Eaat2
CIC.a
axo
CG9657
3 11 5 6 7 9 12 0 2 8 1 4 10
Eaat2
C
wrapper
zyd
Akap200
11 12 5 4 6 8 10 1 3 7 9 0 2
wrapper
D
SPARC
moody
Mdr65
4 5 11 12 0 6 1 9 8 10 7 2 3
Mdr65
UMAP_1
UMAP_2
E
9
7
CORTEX GLIA
14.7%
2
ENSHEATHING GLIA
13.8%
3
1
0
10
8
4
SURFACE GLIA
ASTROCYTES
14.4%
6
11
5
12
UMAP 2
UMAP 1
Undifferentiated 0
Unidentified 1
Cortex Glia 2
Ensheathing Glia 3
Surface Glia 4
Astrocyte 5
Astrocyte 6
Neuroblasts 7
Unidentified 8
Midline Glia 9
Fat Body 10
Astrocyte 11
Astrocyte 12

Figure 4: scRNA seq of larval CNS identified all major glial cell populations

(A-D) Heatmaps (left), feature plots (middle) and Dotplots (right) of known gene markers used for the identification of cell populations in larval 7K data set. Highlighting glial cell clusters of (A) astrocytes, (B) ensheathing glia, (C) cortex glia and (D) surface glia. (E) Hierarchical dendrogram (left) and UMAP representation (right) of clusters in larval data set, with their assigned glial cell type and the proportion (%) of the 7K cells that each cluster represents.

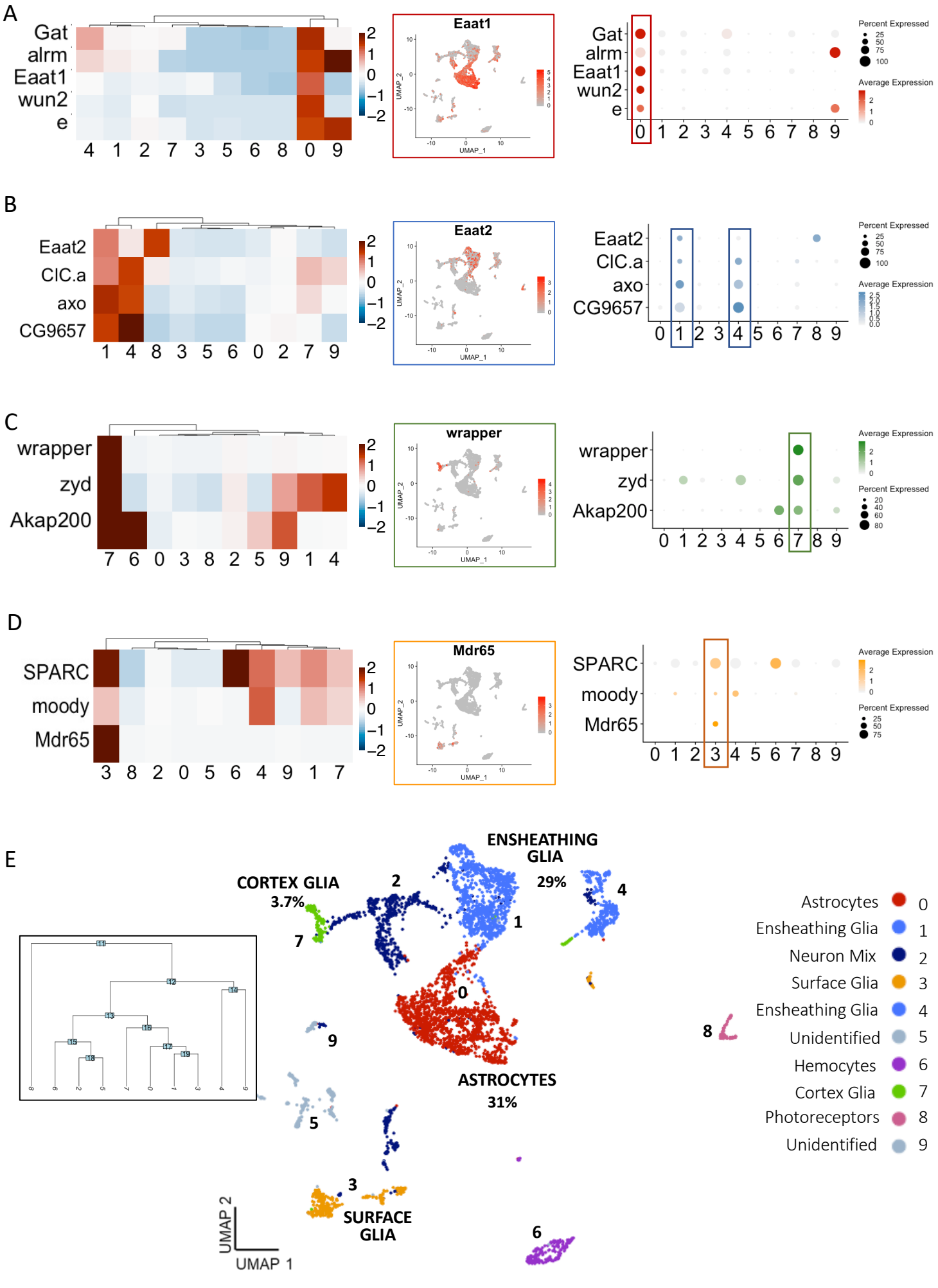
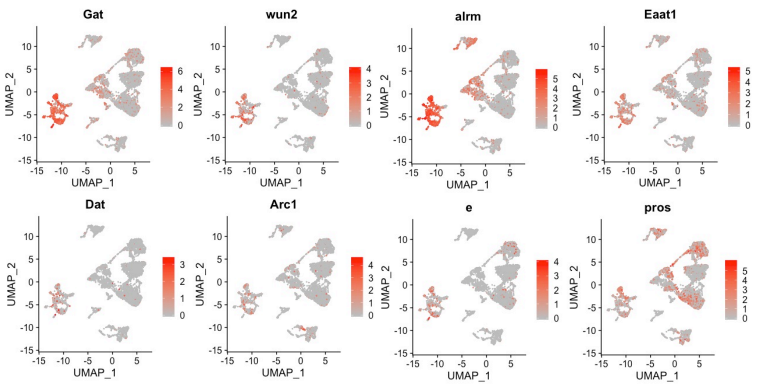

A
Gat
alrm
Eaat1
wun2
e
4 1 2 7 3 5 6 8 0 9
2 1 0 −1 −2
Eaat1
UMAP_1
UMAP_2
0 1 2 3 4 5 6 7 8 9
Percent Expressed
25 50 75 100
Average Expression
B
Eaat2
ClC.a
axo
CG9657
1 4 8 3 5 6 0 2 7 9
Eaat2
0 1 2 3 4 5 6 7 8 9
C
wrapper
zyd
Akap200
7 6 0 3 8 2 5 9 1 4
wrapper
0 1 2 3 4 5 6 7 8 9
D
SPARC
moody
Mdr65
3 8 2 0 5 6 4 9 1 7
Mdr65
SPARC
moody
Mdr65
0 1 2 3 4 5 6 7 8 9
E
ENSHEATHING GLIA
29%
CORTEX GLIA
3.7%
ASTROCYTES
31%
SURFACE GLIA
UMAP 2
UMAP 1
Astrocytes 0
Ensheathing Glia 1
Neuron Mix 2
Surface Glia 3
Ensheathing Glia 4
Unidentified 5
Hemocytes 6
Cortex Glia 7
Photoreceptors 8
Unidentified 9

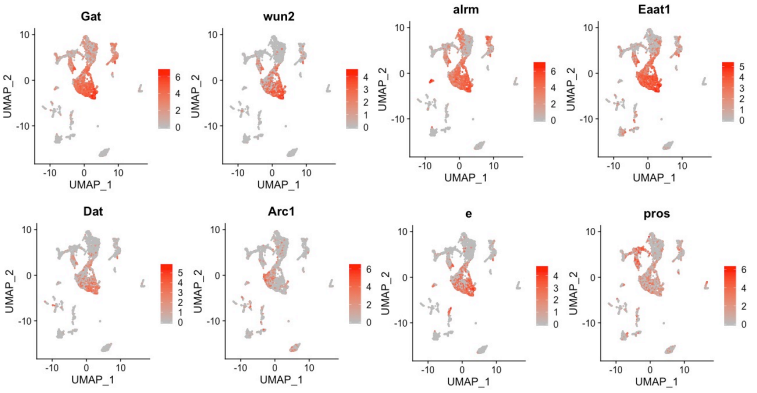
Figure 5: scRNA seq of adult CNS identified all major glial cell populations

(A-D) Heatmaps (left), feature plots (middle) and Dotplots (right) of known gene markers used for the identification of cell populations in adult 4.4K data set. Highlighting glial cell clusters of (A) astrocytes, (B) ensheathing glia, (C) cortex glia and (D) surface glia. (E) Hierarchical dendrogram (left) and UMAP representation (right) of clusters in adult data set, with their assigned glial cell type and the proportion (%) of the 4.4K cells that each cluster represents.

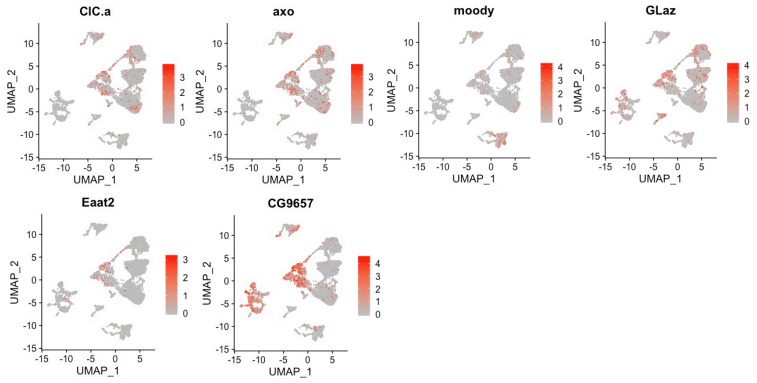
A Astrocyte marker expression in larval 7K



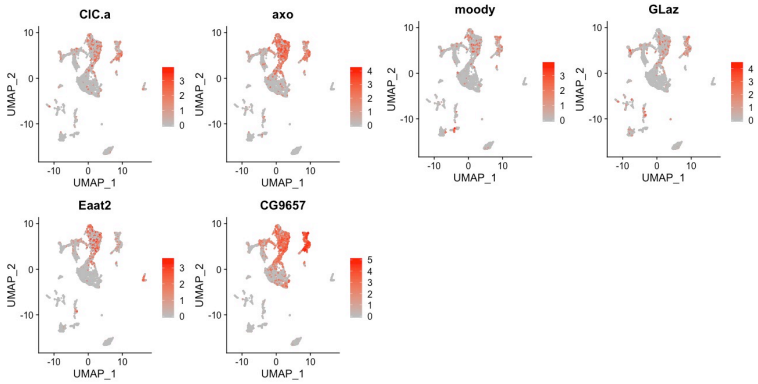
B Astrocyte marker expression in adult 4.4K



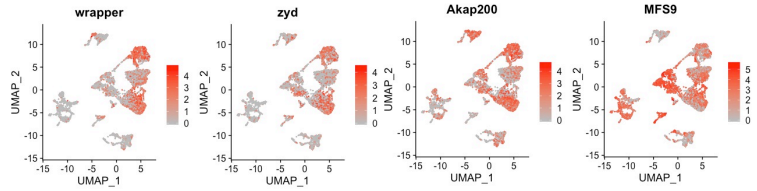
C Ensheathing glia marker expression in larval 7K



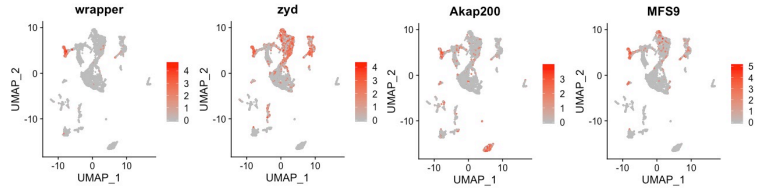
D Ensheathing glia marker expression in adult 4.4K



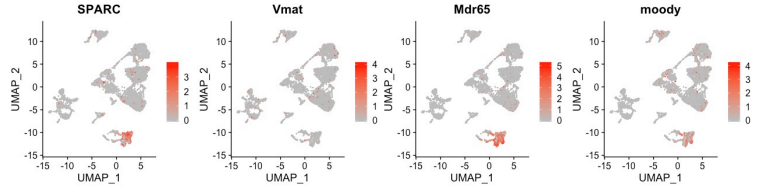
E Cortex glia marker expression in larval 7K



F Cortex glia marker expression in adult 4.4K



G Surface glia marker expression in larval 7K



H Surface glia marker expression in adult 4.4K

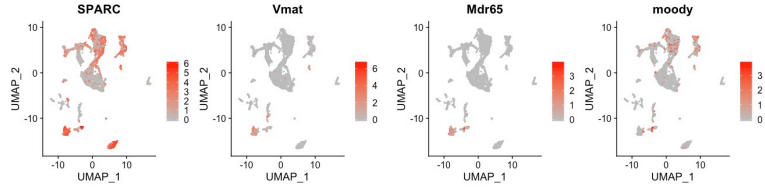
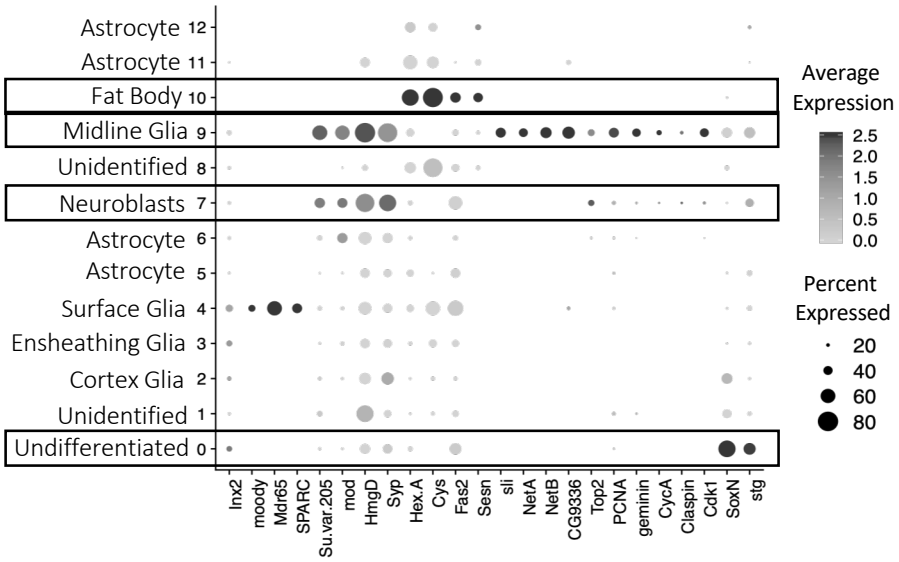


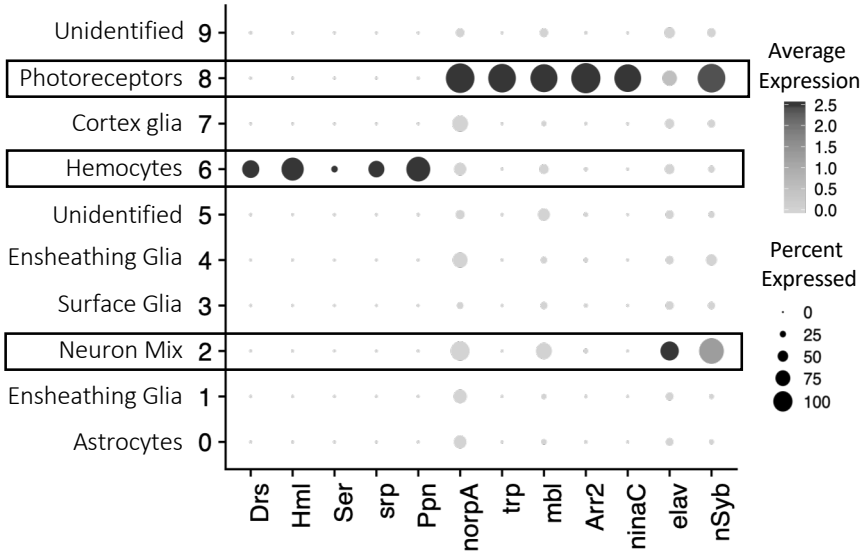
Figure 6: Assignment of identities to glial clusters in Larval 7K and Adult 4.4K data sets using expression of known markers for each cell type.

Feature plots for (left) larval 7K and (right) adult 4.4K data sets, showing expression of known markers of astrocytes (A,B), ensheathing glia (C,D), cortex glia (E,F) and surface glia (G,H). Single cells in the UMAP are shown with color intensity representing relative expression level.

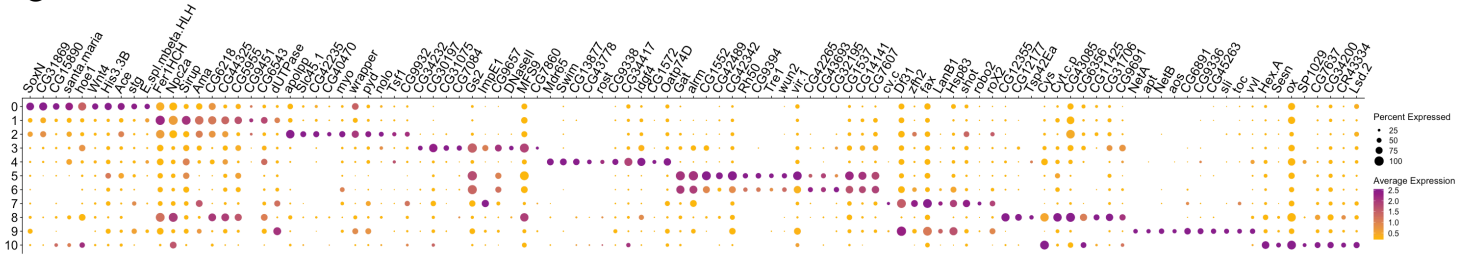
A


DotPlot of gene expression for Larval 7K
Astrocyte 12
Astrocyte 11
Fat Body 10
Midline Glia 9
Unidentified 8
Neuroblasts 7
Astrocyte 6
Astrocyte 5
Surface Glia 4
Ensheathing Glia 3
Cortex Glia 2
Unidentified 1
Undifferentiated 0
lnx2
moody
Mdr65
SPARC
Su.var.205
mod
HmgD
Syp
Hex.A
Cys
Fas2
Sesn
sll
NetA
NetB
CG9336
Top2
PCNA
geminin
CycA
Claspin
Cdk1
SoxN
stg
Average Expression
2.5
2.0
1.5
1.0
0.5
0.0
Percent Expressed
20
40
60
80

B


DotPlot of gene expression for Adult 4.4K
Unidentified 9
Photoreceptors 8
Cortex glia 7
Hemocytes 6
Unidentified 5
Ensheathing Glia 4
Surface Glia 3
Neuron Mix 2
Ensheathing Glia 1
Astrocytes 0
Drs
Hml
Ser
srp
Ppn
norpA
trp
mbl
Arr2
ninaC
elav
nSyb
Average Expression
2.5
2.0
1.5
1.0
0.5
0.0
Percent Expressed
0
25
50
75
100

C


Percent Expressed
Average Expression

D

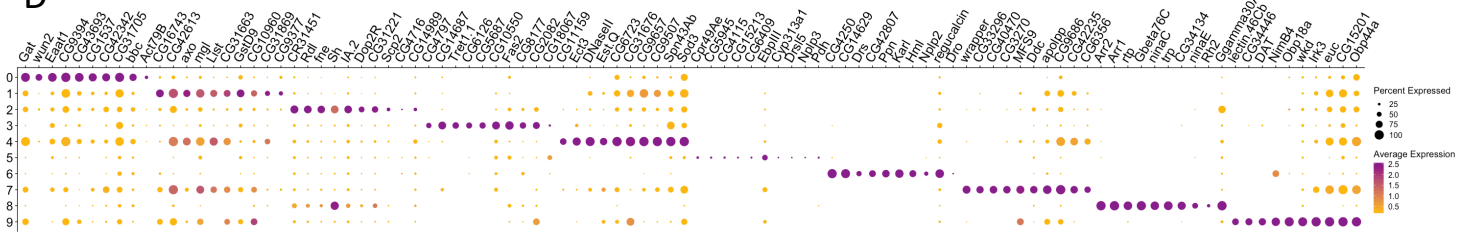
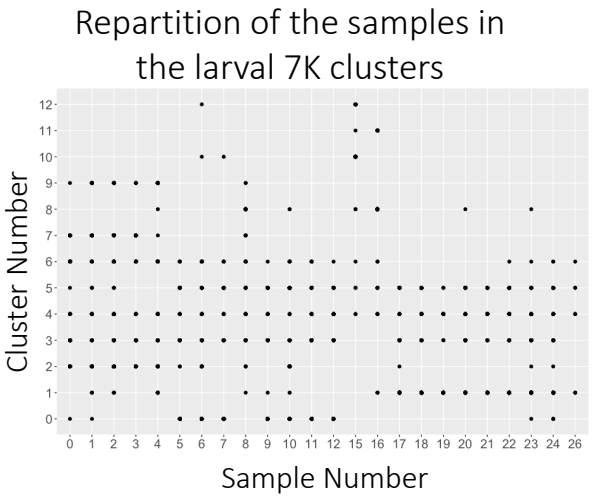

Percent Expressed
Average Expression

Figure 7: scRNAseq of larvae and adults identified other glial and non-glia cell types

Dot plot displaying the expression of known marker genes for other cell clusters in (A) larval 7K and (B) adult 4.4K data sets, where dot size represents percent of cells in cluster expressing gene of interest and the dot color intensity reflects average scaled expression (0-2.5). Clusters identified using these markers are highlighted (box).(C-D) Dot plot of the top 10 DEG markers for each glial sub-type based on fold-enrichment for (C) larval 7K data set, where small astrocyte clusters 11 and 12 were not included due to their redundancy with clusters 5 and 6 and (D) adult 4.4K data set.

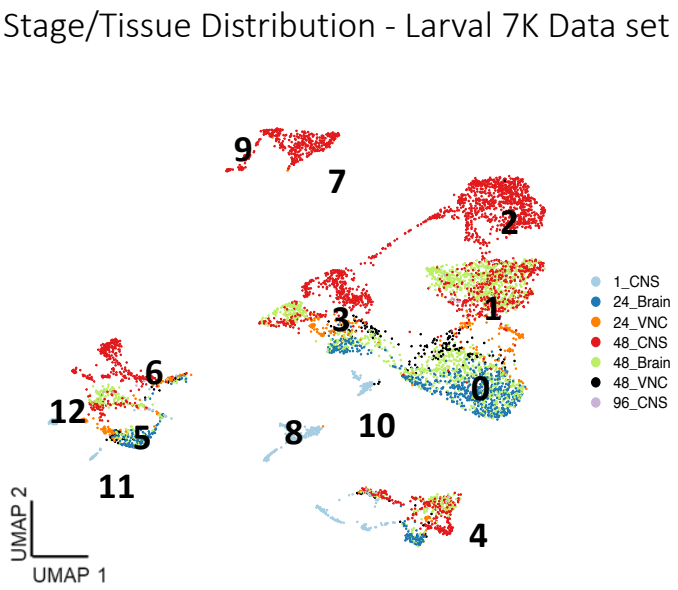
A



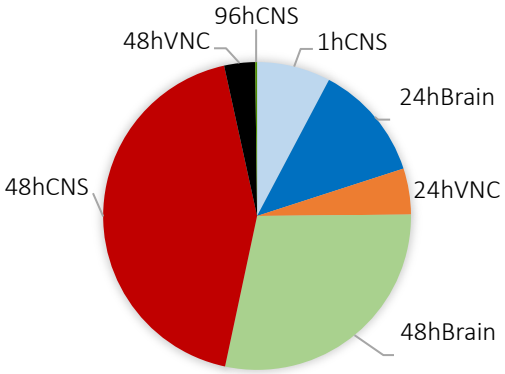
B

Sample Number	0	1	2	3	4	5	6	7
Stage (hours ALH)	48	48	48	48	48	48	48	24
Tissue	1 CNS	1 CNS	2 CNS	1 CNS	1 CNS	6 brains	4 VNC	10 brains
Sample Number	8	9	10	11	12	15	16	17
Stage (hours ALH)	24	24	24	48	24	1	1	48
Tissue	6 VNC	10 brains	6 VNC	8 brains	6 brains	4 CNS	5 CNS	8 brains
Sample Number	18	19	20	21	22	23	24	26
Stage (hours ALH)	48	48	48	48	48	48	48	96
Tissue	8 brains	3 CNS	3 CNS	8 brains	8 brains	3 CNS	3 CNS	6 CNS

C

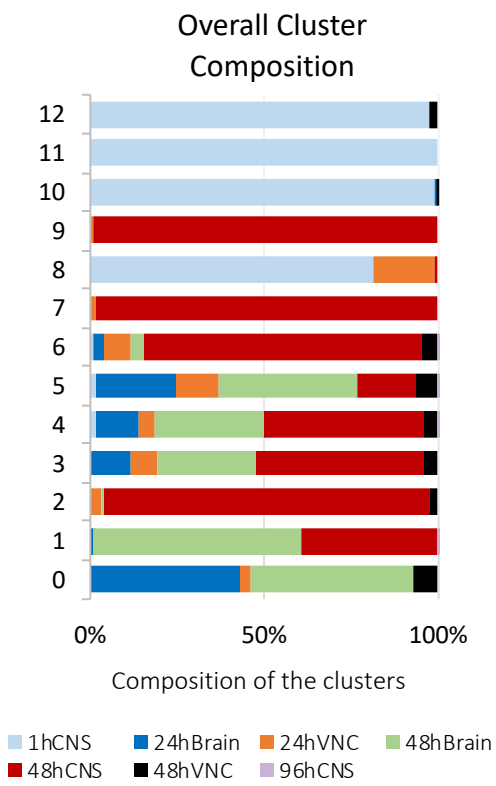


D

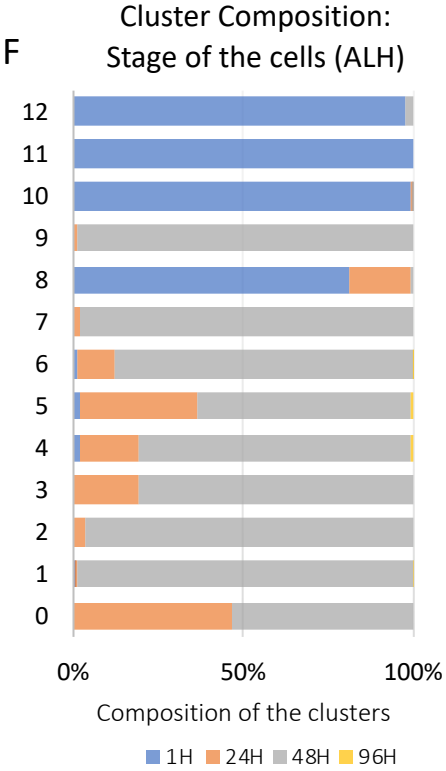


Sample type	Number of cells	%
1h CNS	549	7.7
24h Brain	871	12.3
24h VNC	343	4.8
48h Brain	2021	28.5
48h CNS	3066	43.2
48h VNC	230	3.2
96h CNS	14	0.2

E



F



G

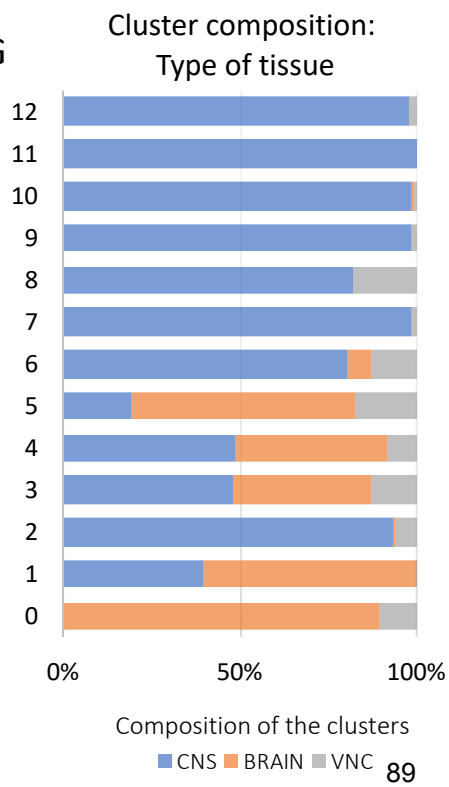
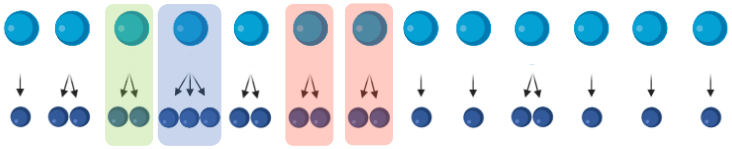


Figure 8: Cluster composition in larval 7K data set

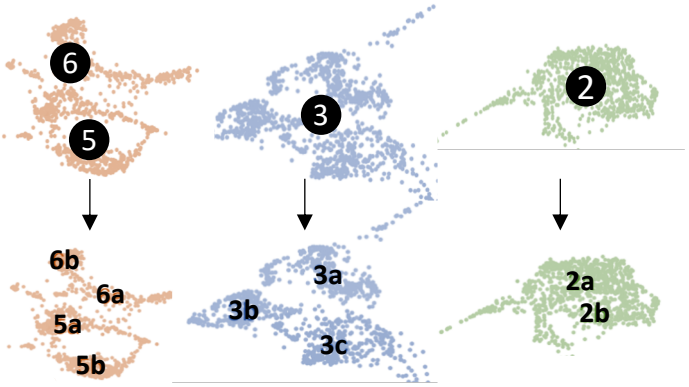
(A) Samples were evenly distributed in the clusters 0-6 , where a dot indicates the presence of that sample in the cluster. (B) Sample distribution among clusters. (C) Distribution of samples (stage and tissue dissection) overlapped onto the larval 7K UMAP. (D) Pie chart and table indicating number and proportion of cells in the data set from specific sample types. (E-G) Summary bar graphs of the proportion of cells in each cluster based on (E) larval age and tissue dissection (F) larval age only and (G) tissue source.

A

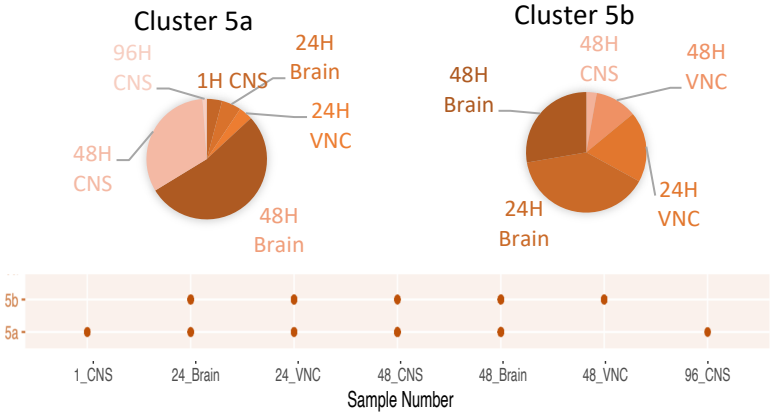
Clustering of larval 7K from Resolution 0.2 to Resolution 0.6

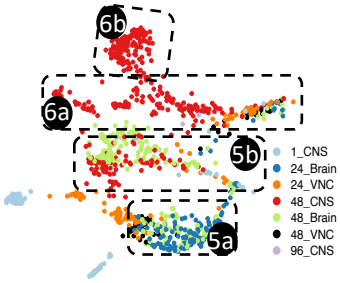


B

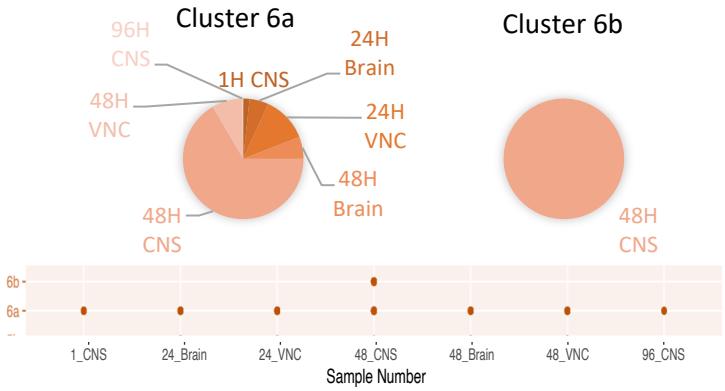


C

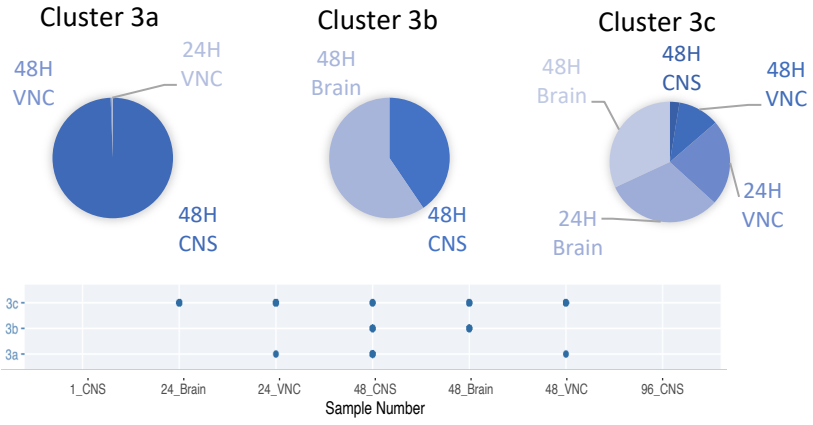


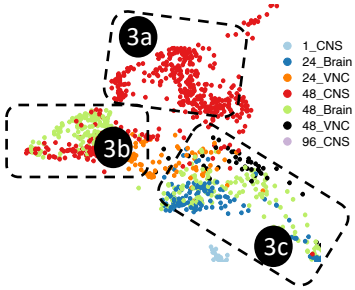


D

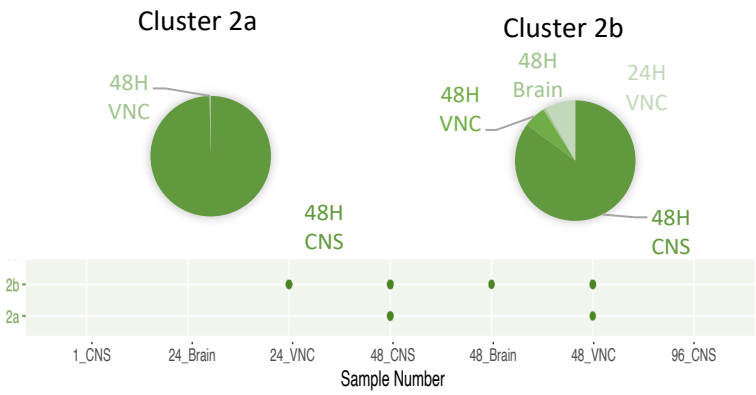


E





F



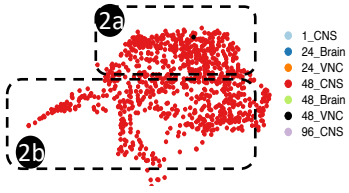


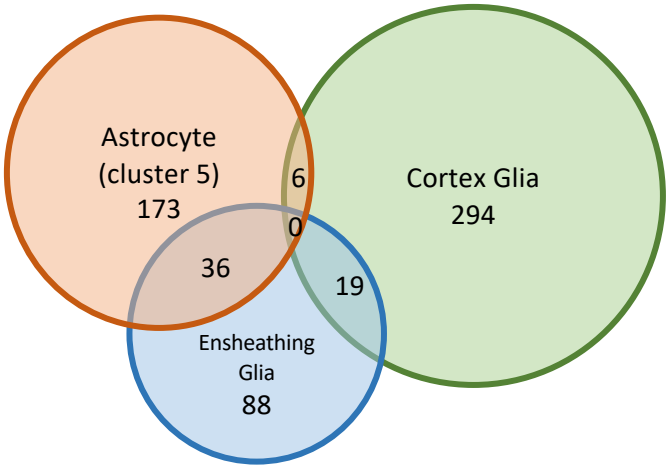
Figure 9: Heterogeneity in larval clusters can be attributed to age and stage of CNS

(A) Diagram of larval 7K data when clustering at resolution 0.2 (top) or resolution 0.6 (bottom). (B) UMAP representation of original clusters (res 0.2) and new sub-clusters (res 0.6) for astrocytes (left), ensheathing glia (middle) and cortex glia (right). Pie charts, dot plots and feature plots of sample distribution within clusters at resolution 0.6 for (C) astrocyte cluster 5, (D) astrocyte cluster 6, (E) ensheathing glia and (F) cortex glia.

A Summary Table of number of cells and markers for favorite clusters

Cell Type	Astrocytes			Ensheathing Glia			Cortex Glia	
Cluster	Larva 5	Larva 6	Adult 0	Larva 3	Adult 1	Adult 4	Larva 2	Adult 7
Number of Cells	554	378	1373	979	951	335	1041	162
Number of Markers	215	324	210	143	367	274	319	382

B Venn Diagram comparing all markers of main glial clusters in Larval 7K data set



C Venn Diagram comparing all markers of main glial clusters in Adult 4.4K data set

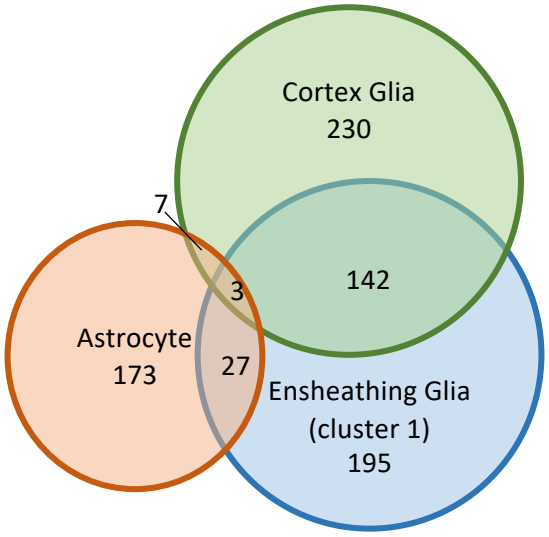
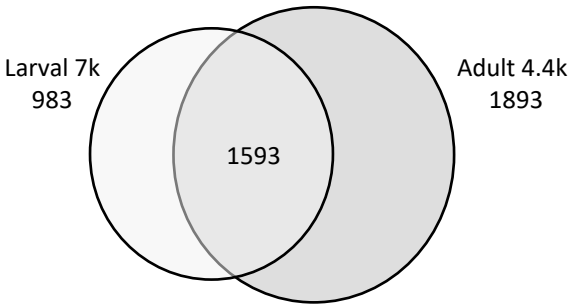


Figure 10. Clusters and markers for astrocytes, ensheathing glia and cortex glia.

(A) Number of cells and number of markers (logfc.threshold = 0.25, test.use = wilcox, min.pct = 0.1, min.diff.pct = -Inf) for main larval 7K and adult 4.4K clusters of astrocytes, ensheathing glia and cortex glia. Venn diagrams show comparison of markers in each cell type for (A) larvae and (B) adult CNS. Numbers of intersecting genes are depicted within the diagrams. For simplicity, only one of the two larval astrocytic clusters (cluster 5 in B) and one of the two adult ensheathing glia clusters (cluster 1 in C) were selected in these representations.

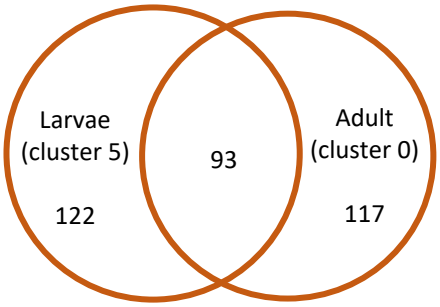
A

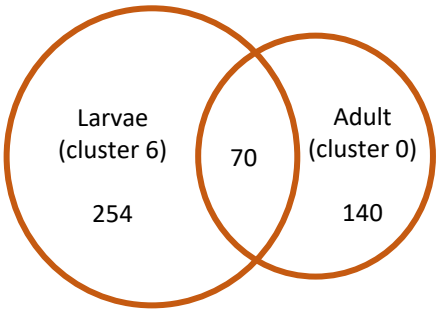
Comparison of all DEGs in data sets



B

Larvae only	Shared	Adult only
CG42489	**alrm**	Act79B
CG11739	CG1537	bbc
CG45057	**CG1552**	CG42741
CG5273	CG42342	CG6465
CG6723	**CG9394**	Chd64
CG9657	**Gat**	CR43459
Galphao	**Gs2**	CR43651
KrT95D	**Rh50**	Msr-110
Pglym78	Tre1	ogre
Vir-1	**wun2**	RpL22



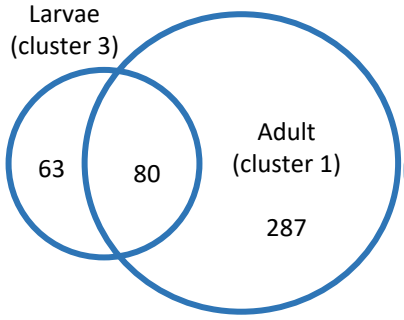


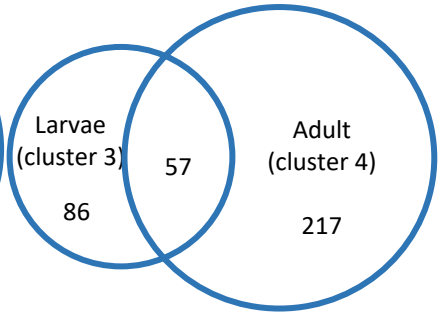
C

Larvae only	Shared	Adult only
Adk1	**alrm**	CG31705
CG10126	**CG1552**	bbc
CG10621	CG42265	CG42741
CG31676	CG43693	Chd64
CG32195	**CG9394**	CR43459
CG8568	**Gat**	flw
CG9657	**Gs2**	ogre
Dr	mbc	pum
Ect3	**Rh50**	rho
Gagr	**wun2**	sty

D

Larvae only	Shared	Adult only
CG1552	**CG7084**	CG10960
CG30197	CAH1	CG3940
CG31075	CG16743	CG5758
CG32444	**CG4829**	CG9507
CG34232	**CG9657**	GstD9
CG43341	**CG9691**	List
CG7860	**DNaseII**	mgl
CG9914	**Elal**	Pkn
Gs2	**ImpE1**	Tsf1
Prip	**MFS9**	Ugt35b

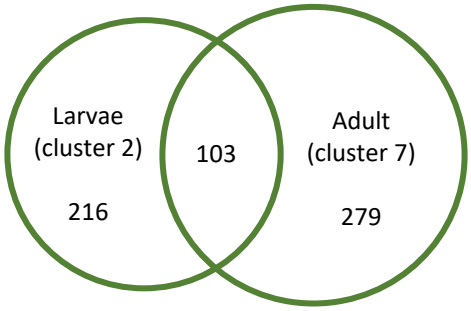




E

Larvae only	Shared	Adult only
CAH1	CG30197	CG11159
CG10361	CG43341	CG31676
CG16743	**CG4829**	CG9507
CG31075	**CG7084**	Ect3
CG32444	**CG9657**	Est-Q
CG34232	**CG9691**	Fas1
CG7860	**DNaseII**	SLC22A
CG9914	**Elal**	Sod3
Gs2	**ImpE1**	Tig
Prip	**MFS9**	Ugt35b

F



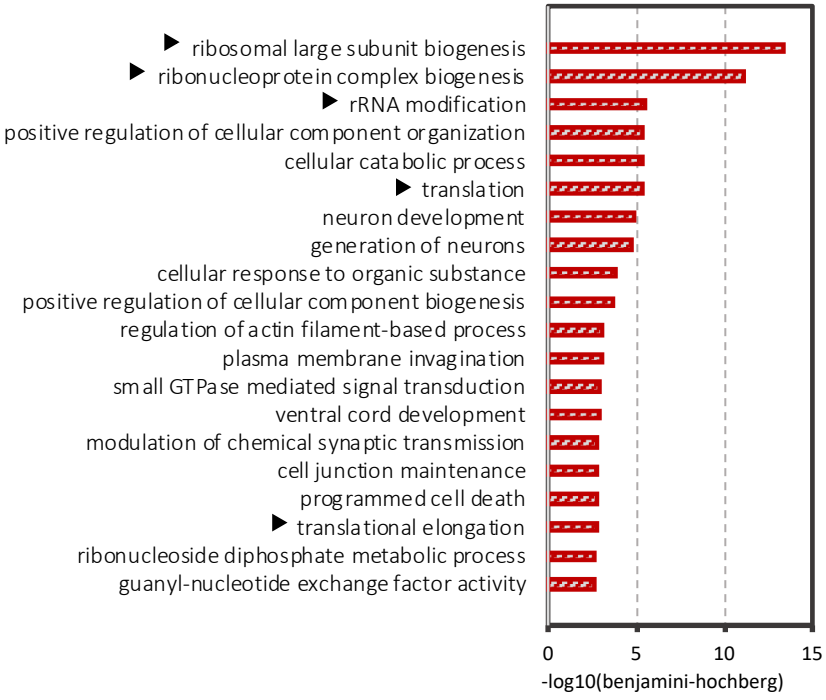
Larvae only	Shared	Adult only
CG43085	apolpp	bru2
CG8839	CG2852	CG14301
Msp300	CG31663	CG33296
mt:lrRNA	CG40470	CG42327
noe	CG42235	Cyp4ac3
Nrt	myo	Dh31-R
Obp99a	Obp44a	Dip-B
path	pyd	Gal
pros	Slc45-1	MFS9
roX2	wrapper	SoxN

Figure 11: Comparison of DEG markers in larvae and adults

Venn diagrams comparing up-regulated DEG markers (logfc.threshold = 0.25, test.use = wilcox, min.pct = 0.1, min.diff.pct = -Inf) for (A) entire larval 7K and adult 4.4K data sets, (B, C) astrocytes (D, E) ensheathing glia and (F) cortex glia. (B-F) lists of the top 10 DEG markers (based on log_FC) in larvae, in adults or shared in both. (B-E) Where there were two clusters for astrocytes (clusters 5 and 6 in larvae) or two clusters of ensheathing glia (clusters 1 and 4 in adults), the genes in bold are those common to both.

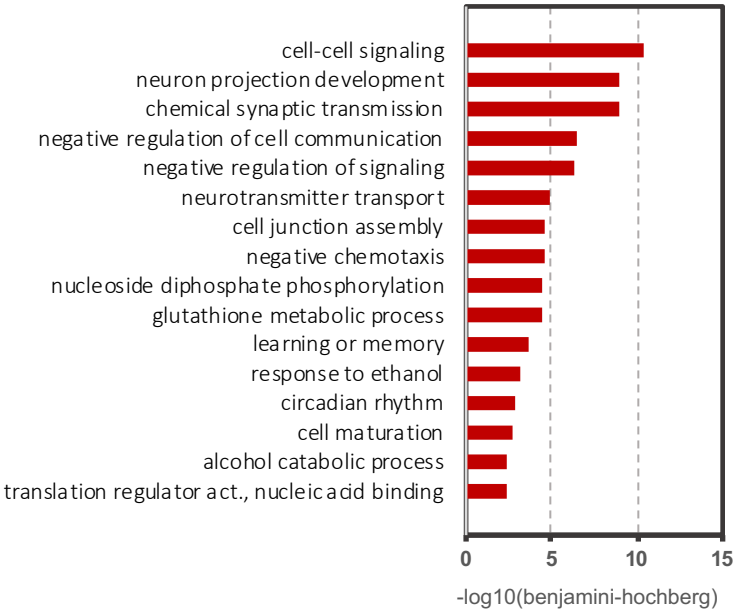
A

Enriched Biological Processes
Larval 7K Astrocytes #6



B

Enriched Biological Processes
Larval 7K Astrocytes #5



C

Enriched Biological Processes
Adult 4.4K Astrocytes

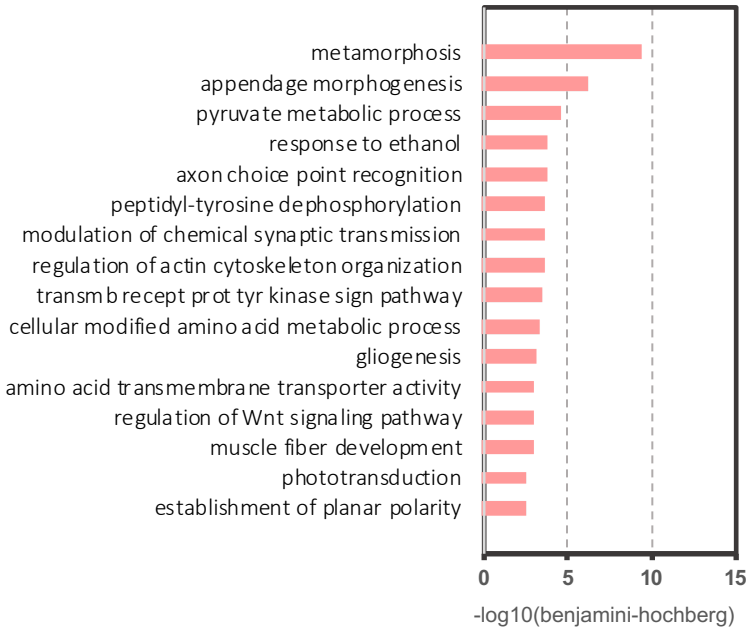
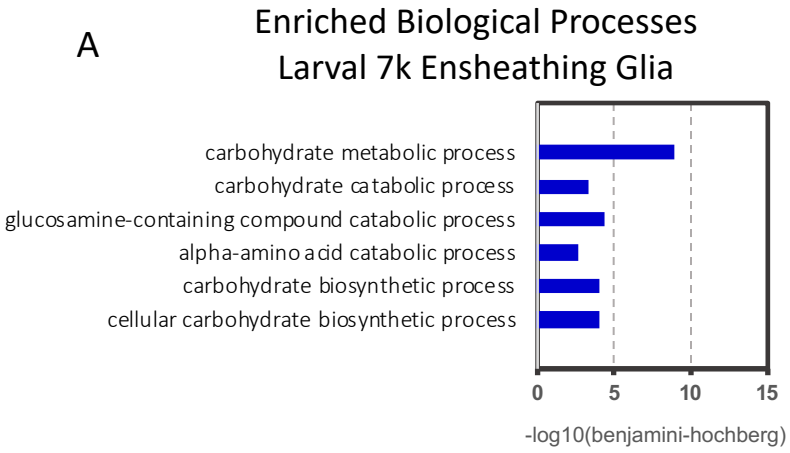


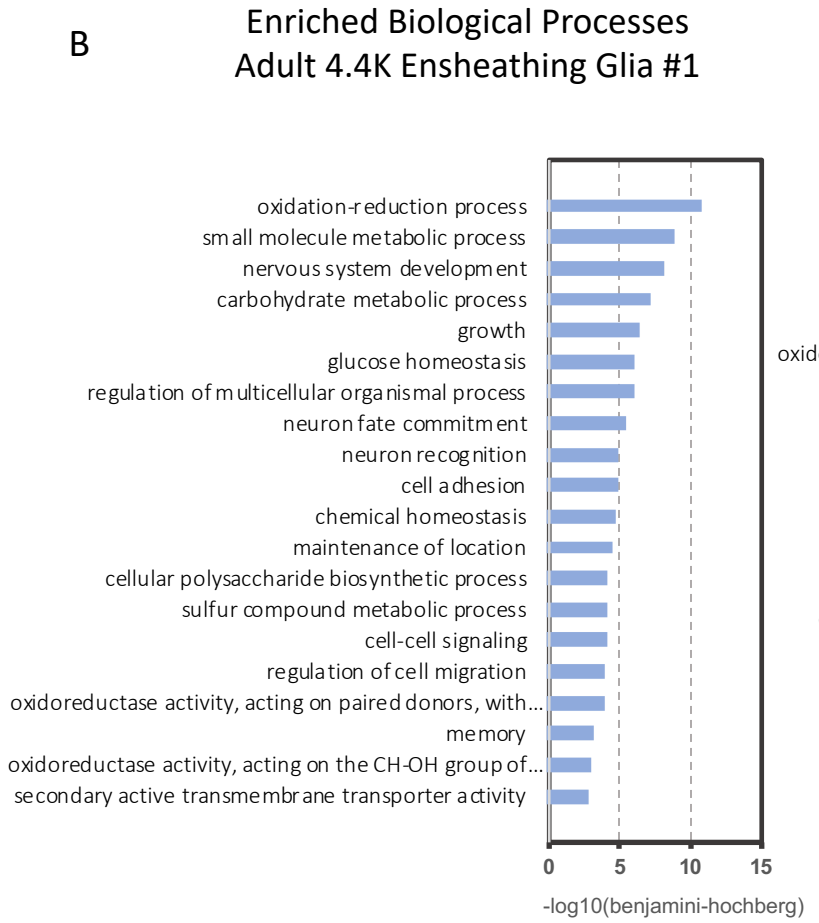
Figure 12: Astrocyte clusters - enriched biological processes from Gene Ontology

Gene ontology analysis using Cytoscape for upregulated biological processes in (A) larval cluster 6, (B) larval cluster 5 and (C) adult cluster 0. (A) Enrichment for factors involved in translation was evident in larval cluster 6 (arrowheads). Figure and analysis in collaboration with Dr. Emilie Peco.

A


Enriched Biological Processes
Larval 7k Ensheathing Glia
carbohydrate metabolic process
carbohydrate catabolic process
glucosamine-containing compound catabolic process
alpha-amino acid catabolic process
carbohydrate biosynthetic process
cellular carbohydrate biosynthetic process
0
5
10
15
-log10(benjamini-hochberg)

B


Enriched Biological Processes
Adult 4.4K Ensheathing Glia #1
oxidation-reduction process
small molecule metabolic process
nervous system development
carbohydrate metabolic process
growth
glucose homeostasis
regulation of multicellular organismal process
neuron fate commitment
neuron recognition
cell adhesion
chemical homeostasis
maintenance of location
cellular polysaccharide biosynthetic process
sulfur compound metabolic process
cell-cell signaling
regulation of cell migration
oxidoreductase activity, acting on paired donors, with...
memory
oxidoreductase activity, acting on the CH-OH group of...
secondary active transmembrane transporter activity
0
5
10
15
-log10(benjamini-hochberg)

C

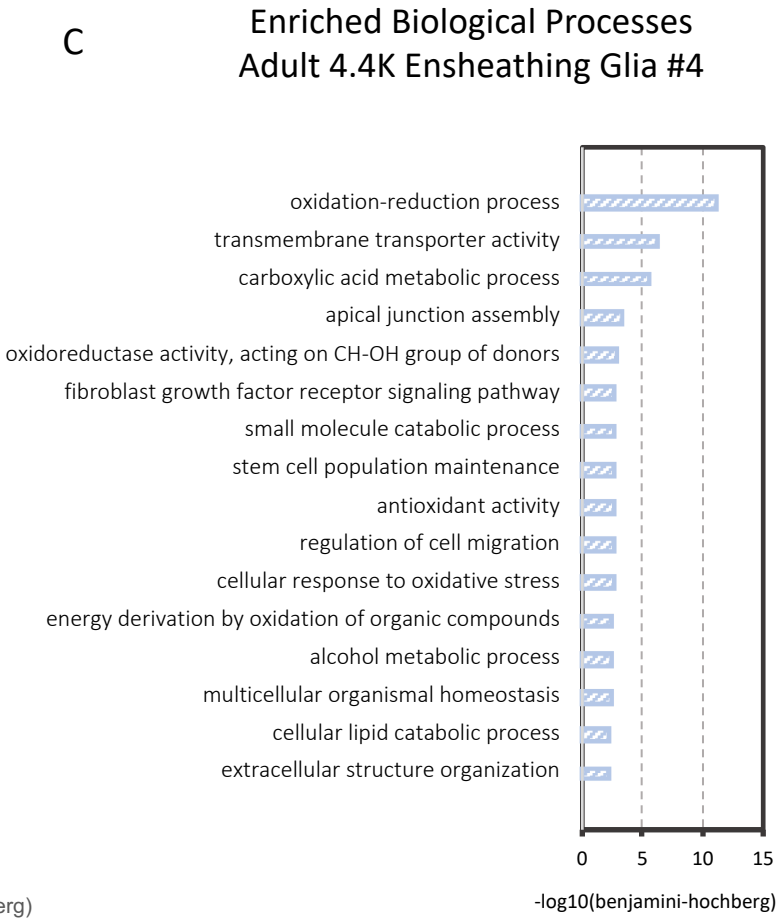
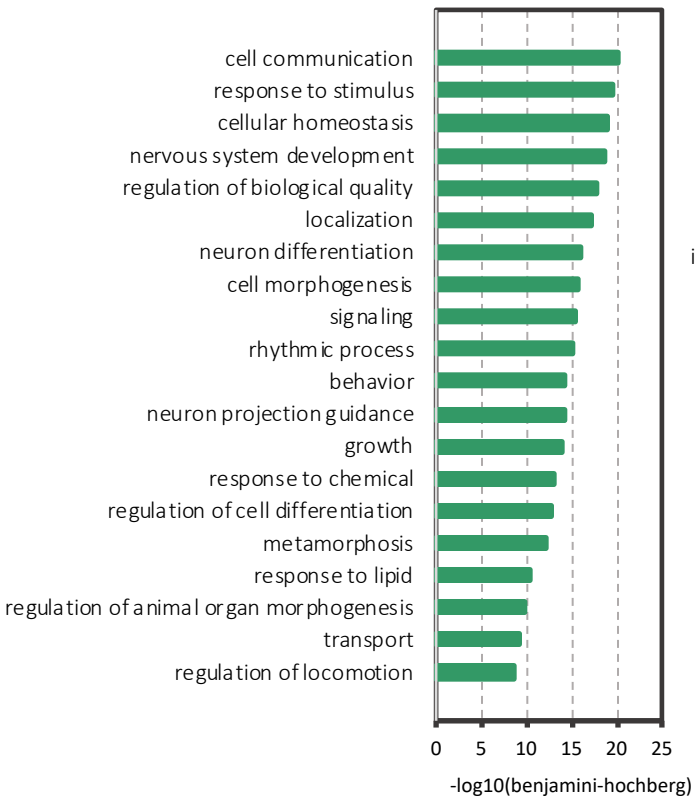

Enriched Biological Processes
Adult 4.4K Ensheathing Glia #4
oxidation-reduction process
transmembrane transporter activity
carboxylic acid metabolic process
apical junction assembly
oxidoreductase activity, acting on CH-OH group of donors
fibroblast growth factor receptor signaling pathway
small molecule catabolic process
stem cell population maintenance
antioxidant activity
regulation of cell migration
cellular response to oxidative stress
energy derivation by oxidation of organic compounds
alcohol metabolic process
multicellular organismal homeostasis
cellular lipid catabolic process
extracellular structure organization
0
5
10
15
-log10(benjamini-hochberg)

Figure 13: Ensheathing Glia clusters - enriched biological processes from Gene Ontology

Gene ontology analysis for upregulated biological processes in (A) larval ensheathing glia cluster 3 (B) adult cluster 1 (C) adult cluster 4. Figure and analysis in collaboration with Dr. Emilie Peco.

A


Enriched Biological Processes
Larval 7K Cortex Glia
cell communication
response to stimulus
cellular homeostasis
nervous system development
regulation of biological quality
localization
neuron differentiation
cell morphogenesis
signaling
rhythmic process
behavior
neuron projection guidance
growth
response to chemical
regulation of cell differentiation
metamorphosis
response to lipid
regulation of animal organ morphogenesis
transport
regulation of locomotion
0
5
10
15
20
25
-log10(benjamini-hochberg)

B

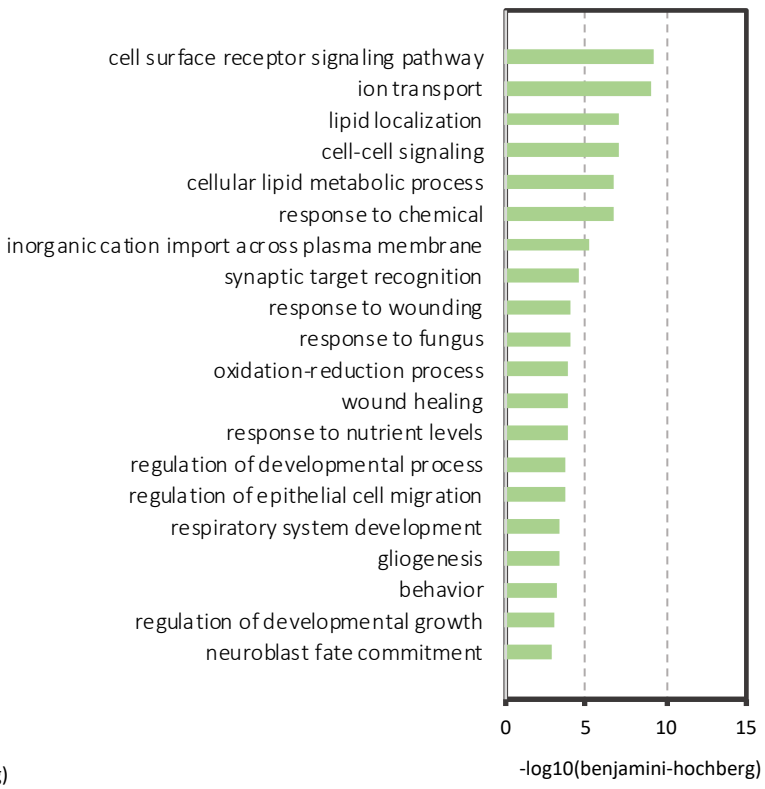
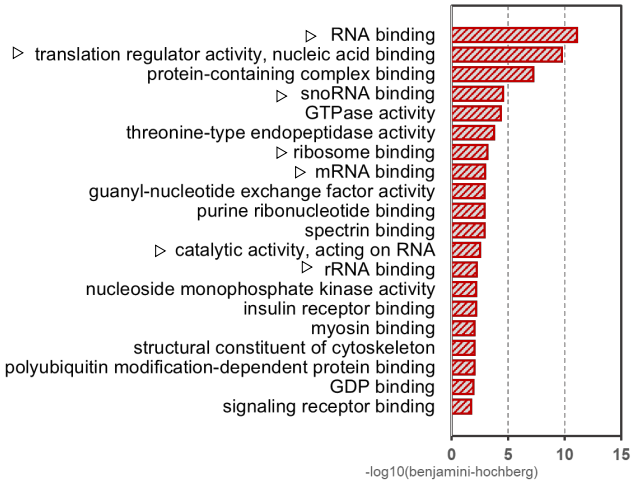

Enriched Biological Processes
Adult 4.4K Cortex Glia
cell surface receptor signaling pathway
ion transport
lipid localization
cell-cell signaling
cellular lipid metabolic process
response to chemical
inorganic cation import across plasma membrane
synaptic target recognition
response to wounding
response to fungus
oxidation-reduction process
wound healing
response to nutrient levels
regulation of developmental process
regulation of epithelial cell migration
respiratory system development
gliogenesis
behavior
regulation of developmental growth
neuroblast fate commitment
0
5
10
15
-log10(benjamini-hochberg)

Figure 14: Cortex Glia clusters - enriched biological processes from Gene Ontology

Gene ontology analysis for upregulated biological processes in cortex glia of the (A) larval 7K and (B) adult 4.4K data sets. Figure and analysis in collaboration with Dr. Emilie Peco.

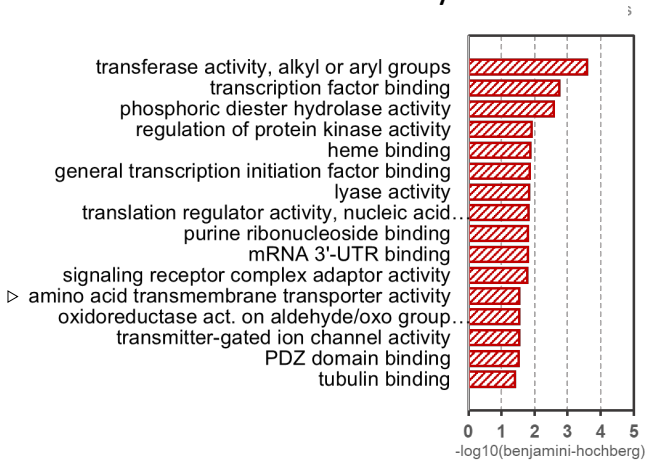
A

Enriched Molecular Functions
Larval 7k Astrocyte #6



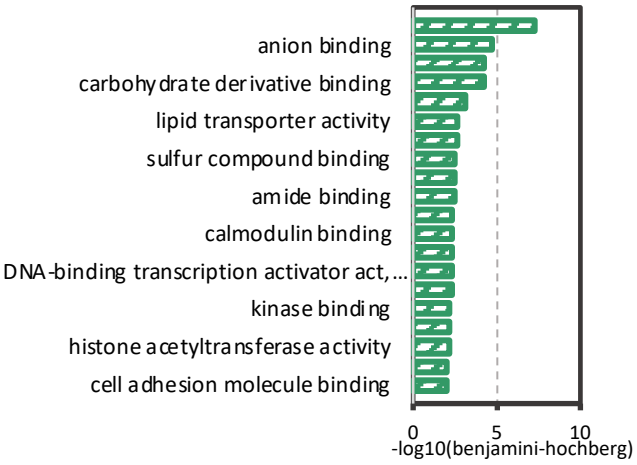
B

Enriched Molecular Functions
Larval 7k Astrocyte #5



C

Enriched Molecular Functions
Larval 7K Cortex Glia



D

Enriched Molecular Functions
Adult 4.4K Cortex Glia

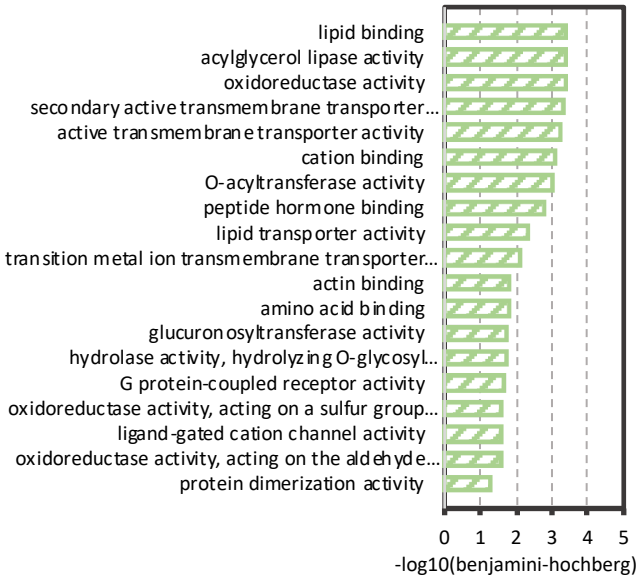


Figure 15: Glial clusters - enriched molecular functions from Gene Ontology

Gene ontology analysis using Cytoscape for upregulated molecular functions in (A) larval astrocyte cluster 6, (B) larval astrocyte cluster 5, (C) larval cortex glia, and D) adult cortex glia. Figure and analysis in collaboration with Dr. Emilie Peco. Further investigation is ongoing, but we have found that Molecular Function categorization of GO terminology was less useful for our purposes than Biological Process.

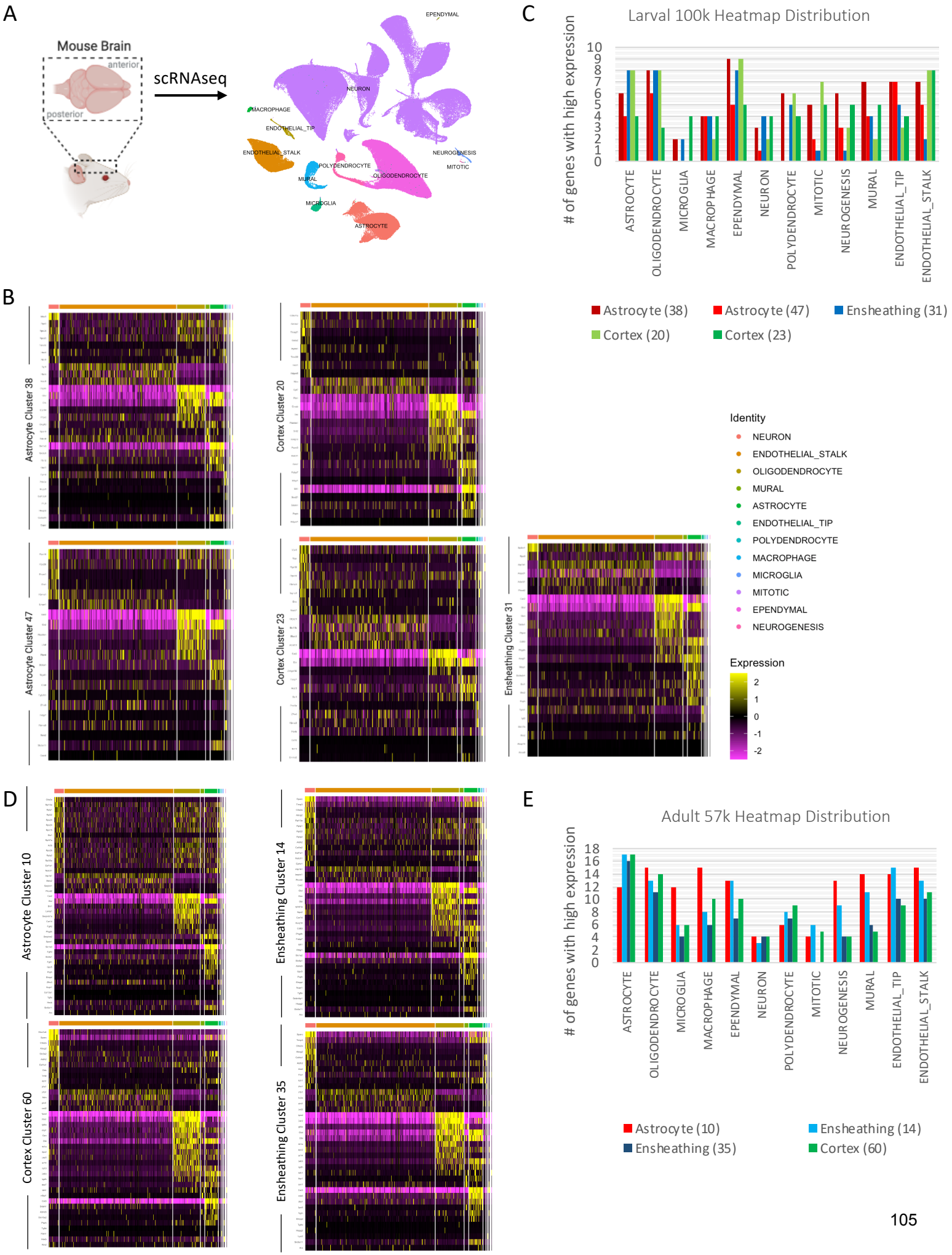

A
Mouse Brain
anterior
posterior
scRNAseq
EPENDYMAL
NEURON
MACROPHAGE
ENDOTHELIAL_TIP
ENDOTHELIAL_STALK
NEUROGENESIS
POLYDENDROCYTE
MITOTIC
MURAL
OLIGODENDROCYTE
MICROGLIA
ASTROCYTE
C
Larval 100k Heatmap Distribution
# of genes with high expression
ASTROCYTE
OLIGODENDROCYTE
MICROGLIA
MACROPHAGE
EPENDYMAL
NEURON
POLYDENDROCYTE
MITOTIC
NEUROGENESIS
MURAL
ENDOTHELIAL_TIP
ENDOTHELIAL_STALK
Astrocyte (38)
Astrocyte (47)
Ensheathing (31)
Cortex (20)
Cortex (23)
B
Astrocyte Cluster 38
Cortex Cluster 20
Astrocyte Cluster 47
Cortex Cluster 23
Ensheathing Cluster 31
Identity
NEURON
ENDOTHELIAL_STALK
OLIGODENDROCYTE
MURAL
ASTROCYTE
ENDOTHELIAL_TIP
POLYDENDROCYTE
MACROPHAGE
MICROGLIA
MITOTIC
EPENDYMAL
NEUROGENESIS
Expression
D
Astrocyte Cluster 10
Ensheathing Cluster 14
Cortex Cluster 60
Ensheathing Cluster 35
E
Adult 57k Heatmap Distribution
# of genes with high expression
ASTROCYTE
OLIGODENDROCYTE
MICROGLIA
MACROPHAGE
EPENDYMAL
NEURON
POLYDENDROCYTE
MITOTIC
NEUROGENESIS
MURAL
ENDOTHELIAL_TIP
ENDOTHELIAL_STALK
Astrocyte (10)
Ensheathing (14)
Ensheathing (35)
Cortex (60)

Figure 16. Expression in the mouse brain of orthologs of DEG markers of Drosophila glia.

(A) Schematic representation of adult mouse brain (left) and cell types identified by the McCarroll lab following scRNAseq (right). (B) Heatmaps showing expression of mouse orthologs of DEG markers from Drosophila larval astrocytes (clusters 38 and 47 from larval 100K data set), larval ensheathing glia (cluster 31) and larval cortex glia (clusters 20 and 23), where expression is shown in 500 randomly selected cells of identified types in the mouse brain (columns: endothelial stalk, neuron, oligodendrocyte, polydendrocytes, astrocytes, mural, endothelial tip, macrophage, microglia, mitotic cells, ependymal cells, neurogenesis). High expression (yellow), low expression (black) and negative expression (purple). (C) Summary bar graph of the number of gene orthologs from Drosophila larvae that are well expressed in mouse brain cell types. (D) Heatmaps showing expression of mouse orthologs of DEG markers from Drosophila adult astrocytes (cluster 10 from adult 57K data set), adult ensheathing glia (clusters 14 and 35), and adult cortex glia (cluster 60), where expression is shown in 500 randomly selected cells of the mouse brain. (E) Summary bar graph of the number of gene orthologs from Drosophila adults that are well expressed in mouse brain cell types. Figure and analysis in collaboration with Dr. Todd Farmer.

Mouse cell populations

Genes from fly glia	Astrocyte	Oligodendrocyte	Microglia	Macrophage	Ependymal	Polydendrocyte	Mitotic	Neurogenesis	Mural	Endoth. stalk	Endoth. tip	Neuron	TOTAL OF HITS
RnrS							1	1					2
Rpl37a				1				1	1	1	1		5
Rpl22			1	1	1			1	1		1		6
RpL35A			1	1	1			1	1	1			6
Rps11			1	1	1			1	1	1			6
Rps20			1	1	1			1	1		1		6
Rplp2			1	1	1			1	1	1	1		7
Rps15		1		1	1			1	1	1	1		7
Rpl13a		1	1	1	1			1	1	1	1		8
Rps14		1	1	1	1			1	1	1	1		8
Rps24		1	1	1	1			1	1	1	1		8
Rps26		1		1	1	1		1	1	1	1		8
Rplp1		1	1	1	1	1		1	1	1	1		9
Rps9		1	1	1	1	1		1	1	1	1		9

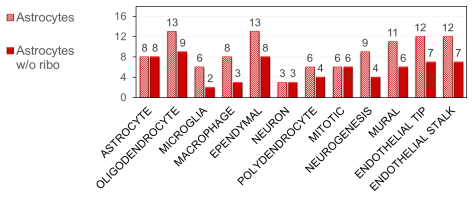
Figure 17: Ribosomal Protein (RP) expression in mouse brain cell populations.

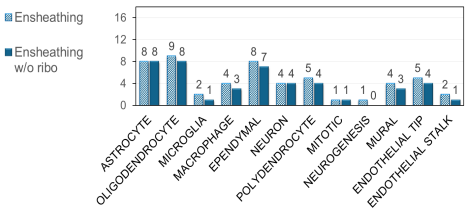
Conserved orthologs of specific RPs that were DEG markers of Drosophila glial cell types are shown to be expressed in cell populations in mouse brain (red), with a sum of hits (in blue).

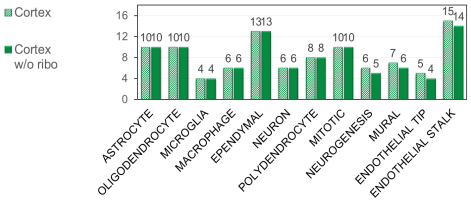
A

Larva 100K	Astrocyte	Oligodendrocyte	Microglia	Macrophage	Ependymal	Neuron	Polydendrocyte	Mitotic	Neurogenesis	Mural	Endothelial_tip	Endothelial_stalk
Astrocyte	8	13	6	8	13	3	6	6	9	11	12	12
Ensheathing	8	9	2	4	8	4	5	1	1	4	5	2
Cortex	10	10	4	6	13	6	8	10	6	7	5	15

Without RPs	Astrocyte	Oligodendrocyte	Microglia	Macrophage	Ependymal	Neuron	Polydendrocyte	Mitotic	Neurogenesis	Mural	Endothelial_tip	Endothelial_stalk
Astrocyte	8	9	2	3	8	3	4	6	4	6	7	7
Ensheathing	8	8	1	3	7	4	4	1	0	3	4	1
Cortex	10	10	4	6	13	6	8	10	5	6	4	14



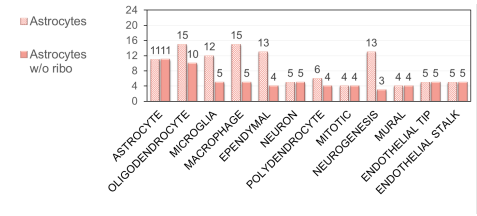


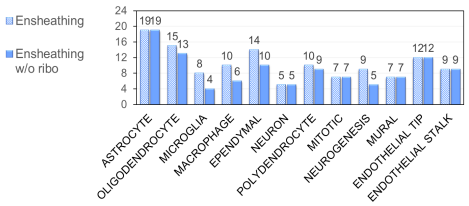


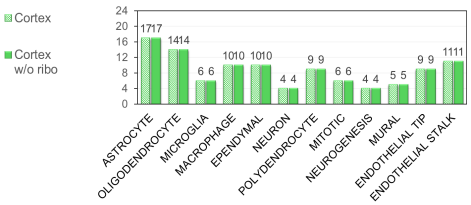
B

Adult 57K	Astrocyte	Oligodendrocyte	Microglia	Macrophage	Ependymal	Neuron	Polydendrocyte	Mitotic	Neurogenesis	Mural	Endothelial_tip	Endothelial_stalk
Astrocyte	11	15	12	15	13	5	6	4	13	14	14	15
Ensheathing	19	15	8	10	14	5	10	7	9	11	16	13
Cortex	17	14	6	10	10	4	9	6	4	5	9	11

Without RPs	Astrocyte	Oligodendrocyte	Microglia	Macrophage	Ependymal	Neuron	Polydendrocyte	Mitotic	Neurogenesis	Mural	Endothelial_tip	Endothelial_stalk
Astrocyte	11	10	5	5	4	5	4	4	3	4	5	5
Ensheathing	19	13	4	6	10	5	9	7	5	7	12	9
Cortex	17	14	6	10	10	4	9	6	4	5	9	11







C

Percent Ribosomal Proteins Present in Overlap

	Mouse											
	Astrocyte	Oligodendrocyte	Polydendrocyte	Endothelial stalk	Endothelial tip	Ependymal	Macrophage	Microglia	Mitotic	Mural	Neurogenesis	Neuron
Larval Astrocytes	0	30	33	42	42	38	63	67	0	45	55	0
Adutl Astrocytes	0	33	33	67	64	69	67	58	0	71	77	0
Larval Ensheathing glia	0	10	20	50	12	25	25	50	0	25	100	0
Adult Esheathing glia	0	13	10	30	25	28	40	50	0	36	44	0
Larval Cortex glia	0	0	0	0.7	20	0	0	0	0	14	17	0
Adult Cortex glia	0	0	0	0	0	0	0	0	0	0	0	0

Figure 18: Influence of RPs on comparisons between Drosophila and mouse scRNAseq data sets

Tables showing the number of orthologs expressed in mouse brain cell types (top) when RPs are included or excluded, using orthologs from the entire (A) larval 100k, or (B) adult (57K) data sets. The numbers of orthologs expressed in some mouse brain cell types were unaffected by the exclusion (astrocytes, mitotic cells and neurons), but the numbers in other cell type were. (C) Summary table of the percent RP present in the mouse cell types from each Drosophila glial cell type. We chose to exclude RPs from our subsequent comparisons of orthologous gene expression in Drosophila and mouse glia.

A

GENES	Fly Astro	Fly Ensh.	Fly Cortex	Astrocyte	Oligodendrocyte	Microglia	Macrophage	Ependymal	Polydendrocyte	Mitotic	Neurogenesis	Mural	Endoth. stalk	Endoth. tip	Neuron	TOTAL OF HITS
CG8745	A	A		1												1
dally	B	A		1												1
aay	A		B	1				1								2
Eaat1	B			1				1								2
Gat	B	A		1						1						2
CG31650	L			1				1								2
CG6287	A	A	A	1	1											2
N	A		A	1						1			1			3
cathD	L			1		1	1									3
Mgstl	L			1				1						1		3
Gs2	B	B		1	1									1		3
htl	A			1	1			1						1		4
Lamp1	A		A	1	1		1		1							4
Acbp2	L		B	1	1			1	1	1	1					6
Spn31A	A			1	1	1		1	1	1		1			1	8

B

CAH1	B	B	L		1											1
CAH7	B		A		1											1
CG14141/ Lar	L				1											1
CG4250	L				1											1
Amph	A				1	1										2
CG43324	L				1					1						2
Hsp27	L				1			1								2
CG11000	L		L		1			1								2
CG6287	A	A	A	1	1											2
Gs2	B	B		1	1									1		3
htl	A			1	1			1						1		4
Lamp1	A		A	1	1		1		1							4
eEF1alpha1	A	A	A		1						1	1	1			4
Act79B	A				1	1	1		1		1	1	1			7
Acbp2	L		B	1	1			1	1	1	1					6
Spn31A	A			1	1	1		1	1	1		1			1	8

C

GENES	Fly Astro	Fly Ensh.	Fly Cortex	Astrocyte	Oligodendrocyte	Microglia	Macrophage	Ependymal	Polydendrocyte	Mitotic	Neurogenesis	Mural	Endoth. stalk	Endoth. tip	Neuron	TOTAL OF HITS
CG8745	A	A		1												1
dally	B	A		1												1
CG4598		L	L	1												1
Eaat2		A		1												1
Hydr1		A	A	1												1
Gat	B	A		1						1						2
CG7860		L		1					1							2
CG6287	A	A	A	1	1											2
msi		L		1				1								2
Prx2540-1		A		1				1								2
sprt		L		1				1								2
Gs2	B	B		1	1									1		3
Cys		A	A	1		1	1									3
fabp		A	B	1					1	1	1					4
Aldh		A	A	1			1					1	1	1		5
Timp		A		1				1				1	1	1		5
Irk3		A	B	1	1			1	1	1						5
Npc2a		A	A	1	1	1			1		1					5
SPARC		A	A	1		1		1				1	1	1		6
Tsp42Ee		B	A	1	1		1	1	1					1		6

D

CAH1	B	B	L		1											1
CG14141/CG7607		L			1											1
CG6287	A	A	A	1	1											2
GLaz		A	A		1									1		2
Spn43Ab		A	A		1										1	2
Gs2	B	B		1	1									1		3
Ugt49B1		A	A		1				1							2
Taldo		L			1				1							2
eEF1alpha1	A	A	A		1						1	1	1			4
Gel		B	A		1			1				1		1		4
Irk3		A	B	1	1			1	1	1						5
Npc2a		A	A	1	1	1			1		1					5
Tsp42Ee		B	A	1	1		1	1	1					1		6

E

GENES	Fly Astro	Fly Ensh.	Fly Cortex	Astrocyte	Oligodendrocyte	Microglia	Macrophage	Ependymal	Polydendrocyte	Mitotic	Neurogenesis	Mural	Endoth. stalk	Endoth. tip	Neuron	TOTAL OF HITS
CG1764			B	1												1
CG4598		L	L	1												1
Hydr1		A	A	1												1
Iap1 (?)			A	1												1
aay	A		B	1				1								2
Acsl			L	1					1							2
CG6287	A	A	A	1	1											2
MtnA			L	1				1								2
Dip-B			A	1	1											2
yin			A	1				1								2
N	A		A	1						1			1			3
CD98hc			B	1									1	1		3
Cys		A	A	1		1	1									3
fabp		A	B	1					1	1	1					4
Aldh		A	A	1			1					1	1	1		5
SoxN			L	1				1	1	1	1					5
Lamp1	A		A	1	1			1		1						4
Tsp42Ea			A	1				1	1			1		1		5
Irk3		A	B	1	1				1	1	1					5
Npc2a		A	A	1	1	1			1		1					5
Acbp2	L		B	1	1			1	1	1	1					6
SPARC		A	A	1		1		1				1	1	1		6
Tsp42Ee		B	A	1	1		1	1	1					1		6

F

CAH1	B	B	L		1											1
CAH7	B		A		1											1
CG43658			L		1											1
CHES-1-like			L		1											1
CG11000	L		L		1			1								2
CG6287	A	A	A	1	1											2
GLaz		A	A		1									1		2
Dip-B			A	1	1											2
Spn43Ab		A	A		1										1	2
Ugt49B1		A	A		1				1							2
Hsp23			L		1			1								2
RabX1			L		1					1						2
Sirt2			L		1				1							2
Fer1HCH			B		1		1	1								3
Lamp1	A		A	1	1		1		1							4
eEF1alpha1	A	A	A		1						1	1	1			4
Gel		B	A		1			1				1		1		4
Irk3		A	B	1	1			1	1	1						5
Npc2a		A	A	1	1	1			1		1					5
Acbp2	L		B	1	1			1	1	1	1					6
Tsp42Ee		B	A	1	1		1	1	1					1		6

Figure 19: Expression of orthologs of DEG markers in mouse brain cell types.

(A,B) Orthologs of DEG markers of Drosophila astrocytes that are well expressed in mouse brain cell types with a focus on (A) mouse astrocytes and (B) mouse oligodendrocytes. (C,D) Orthologs of DEG markers of ensheathing glia, with a focus on (C) mouse astrocytes and (D) mouse oligodendrocytes. (E,F) Orthologs of DEG markers of cortex glia, with a focus on (E) mouse astrocytes and (F) mouse oligodendrocytes. Specification on whether the larval (L), adult (A) or both (B) data sets contributed to the gene. Each instance where the ortholog is expressed is in red, and a sum of hits is in blue.

A

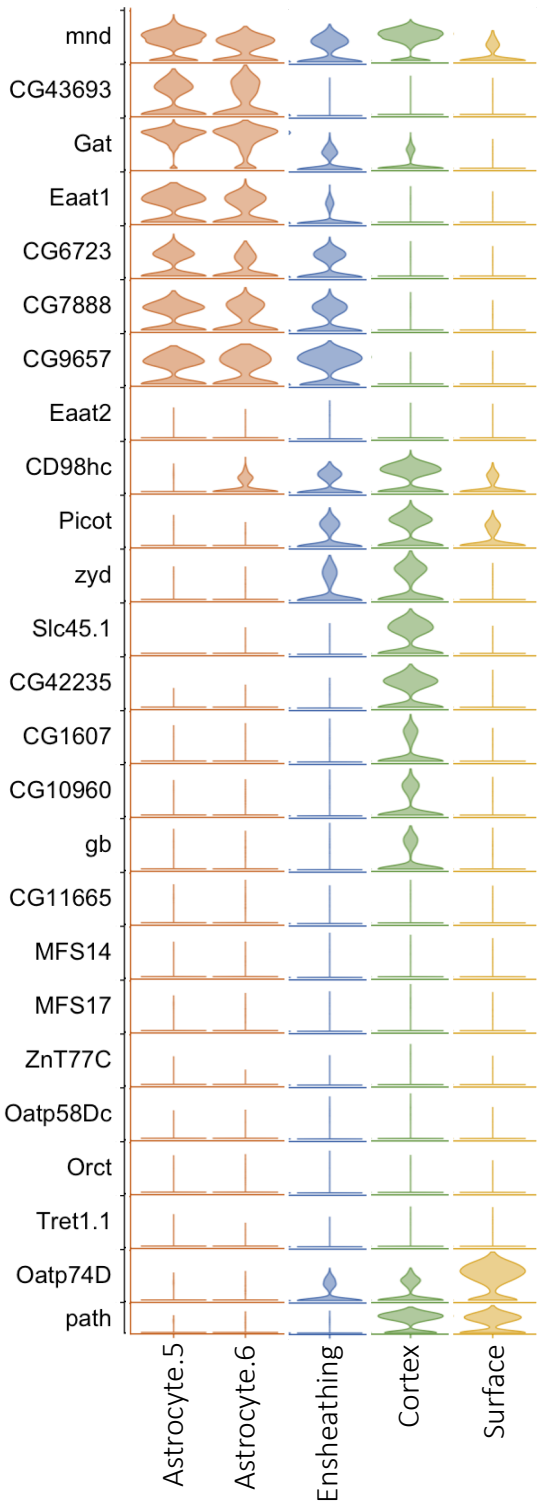

mnd
CG43693
Gat
Eaat1
CG6723
CG7888
CG9657
Eaat2
CD98hc
Picot
zyd
Slc45.1
CG42235
CG1607
CG10960
gb
CG11665
MFS14
MFS17
ZnT77C
Oatp58Dc
Orct
Tret1.1
Oatp74D
path
Astrocyte.5
Astrocyte.6
Ensheathing
Cortex
Surface

Figure 20: Solute Carrier (SLC) transporters in larval glial cells

Stacked violin plots of the SLCs present in the larval 7K data set, arranged by cluster for astrocytes (red), ensheathing glia (blue), cortex glia (green) and surface glia (yellow).

A

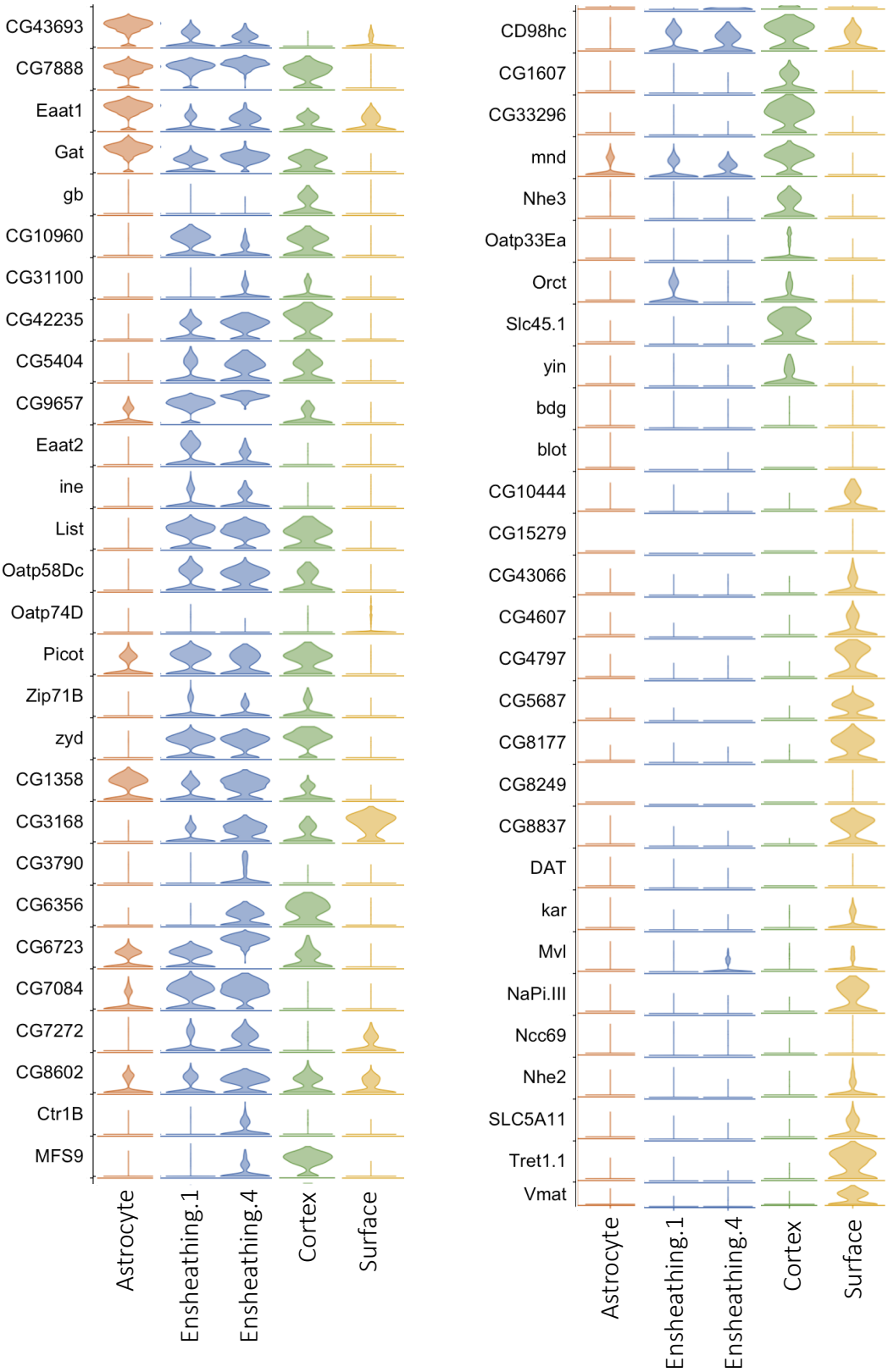

CG43693
CG7888
Eaat1
Gat
gb
CG10960
CG31100
CG42235
CG5404
CG9657
Eaat2
ine
List
Oatp58Dc
Oatp74D
Picot
Zip71B
zyd
CG1358
CG3168
CG3790
CG6356
CG6723
CG7084
CG7272
CG8602
Ctr1B
MFS9
Astrocyte
Ensheathing.1
Ensheathing.4
Cortex
Surface
CD98hc
CG1607
CG33296
mnd
Nhe3
Oatp33Ea
Orct
Slc45.1
yin
bdg
blot
CG10444
CG15279
CG43066
CG4607
CG4797
CG5687
CG8177
CG8249
CG8837
DAT
kar
Mvl
NaPi.III
Ncc69
Nhe2
SLC5A11
Tret1.1
Vmat
Astrocyte
Ensheathing.1
Ensheathing.4
Cortex
Surface

Figure 21: SLC transporters in adult glial cells

Stacked violin plots of the SLCs present in the adult 4.4K data set, arranged by cluster for astrocytes (red), ensheathing glia (blue), cortex glia (green) and surface glia (yellow).