DROSOPHILA MIDLINE DEVELOPMENT: THE ROLE OF *18-WHEELER* AND AN OPTIMIZED PROTOCOL FOR TRANSCRIPTOME ANALYSIS

AMANDA CHRISTINE MOAWAD

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Curriculum in Neurobiology in the School of Medicine.

Chapel Hill
2016

Approved by:

Stephen T. Crews

Richard E. Cheney

Stephen L. Rogers

ABSTRACT

AMANDA CHRISTINE MOAWAD: *Drosophila* Midline Development: The Role of *18-Wheeler* and an Optimized Protocol for Transcriptome Analysis
(Under the direction of Stephen T. Crews)

The developing midline of *Drosophila* consists of diverse cell types that must migrate and differentiate appropriately to form a functional central nervous system. Despite the paucity of midline cells, much is unknown about how these cells migrate to their final locations and acquire fates. In a previous RNA-sequencing screen, we found that *18-Wheeler* (*18w*), a Toll-family receptor, is present in the midline during embryonic development. Fluorescence *in situ* hybridizations and genetic analyses revealed that *18w* is expressed posteriorly in segments in the median neuroblast. *18w* overexpression led to midline disorganization, while heterozygous *18w* loss of function mutants showed no phenotype, indicating that one wild-type copy of *18w* is sufficient to maintain function.

To further analyze midline gene expression, a protocol was optimized to dissociate cells from tissue for fluorescence-activated cell sorting. Isolated midline cells would subsequently undergo single-cell RNA-sequencing, revealing expression differences among cell types and allowing for transcriptome analysis.

To my parents, who tolerated my babbling about fruit flies for years.
Thank you for all your love and support.

ACKNOWLEDGMENTS

This project would not have been possible without the support of many people. Many thanks to my advisor Steve Crews, who provided guidance and pushed me to become an independent scientist. I have been fortunate to have a supportive, helpful and expert committee, and I thank Richard Cheney and Steve Rogers for that. I also thank Stephanie Gupton and Bob Duronio for all of their support and suggestions. I am lucky to have worked with excellent postdocs in the lab, and I thank Joe, Nicole, and Yizhuo for all of their advice, from flies to experimental design to surviving graduate school. Many thanks to Garret Stuber, who has gone above and beyond to help me make this thesis possible. Thanks to my friends, especially Nick and Jennifer, for their moral support and insightful discussions throughout the years. Thanks to Trevor for helpful edits and always making me smile. Lastly, I would like to thank my parents and brother for their love, support, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

18w	*18-wheeler*
AEL	after egg laying
AMG	anterior midline glia
AMP	antimicrobial peptide
BDGP	Berkeley *Drosophila* Genome Project
β-gal	beta-galactosidase
Ca^{2+}	calcium ion
cDNA	complementary DNA
CNS	central nervous system
CRISPR	clustered regularly interspaced short palindromic repeats
CyO	Curly of Oster
DAPI	4',6-diamidino-2-phenylindole
Dif	Dorsal-related immunity factor
DIG	digoxigenin
DNA	deoxyribonucleic acid
dToll	*Drosophila Toll*
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced GFP
elav	*embryonic lethal abnormal vision*
EMS	ethyl methanesulfonate
en	*engrailed*
FACS	fluorescence-activated cell sorting

FISH	fluorescent *in situ hybridization*
ftz	fushi terazu
gad1	*glutamic acid decarboxylase 1*
gDNA	genomic DNA
GFP	green fluorescent protein
gsb	*gooseberry*
HL	hemolymph-like
IL-1R	interleukin-1 receptor
LCM	laser capture microdissection
LOF	loss of function
LRR	leucine rich repeat
mas	*masquerade*
Mg^{2+}	magnesium ion
MG	midline glia
MNB	median neuroblast
MP	midline precursor neuron
mRNA	messenger RNA
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NSC	neural stem cell
odd	*odd-skipped*
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PI	propidium iodide

ple	*pale*
PMG	posterior midline glia
PRR	pattern recognition receptor
P-Sqh	phospho-Spaghetti squash
RNA	ribonucleic acid
RNA-seq	RNA sequencing
scRNA-seq	single cell RNA sequencing
SHH	sonic hedgehog
sim	*single-minded*
Spz	Spätzle
STG	*sim-Gal4; UAS-tau-GFP*
TIR	Toll/IL-1R, refers to a homologous structural domain in TLRs
Tl	*Toll*
TLR	Toll-like receptor
tup	*tailup*
VGlut	*vesicular glutamate transporter*
VUM	ventral unpaired median neuron
WNT	Wingless-related integration site
wor	*worniu*
zfh1	*zinc finger homeodomain 1*

CHAPTER ONE: *18-WHEELER* AFFECTS EARLY MIDLINE DEVELOPMENT

INTRODUCTION

Cell migration during development

Recent estimates suggest that the human brain contains a staggering 86 billion neurons, with an approximately equal number of non-neuronal cells such as glia (Azevedo et al., 2009). During embryogenesis, these cells must migrate to their appropriate locations in the brain, organize, and form connections in order to produce a functioning nervous system. Cell migration is a highly conserved, critical mechanism despite variations in cell types and lineages. Perturbing this elegantly choreographed system leads to aberrant cell migration, which has implications in cancer metastasis and some neurodevelopmental disorders (Casanova, 2014; Kurosaka & Kashina, 2008; Muraki & Tanigaki, 2015; Yamaguchi et al., 2005).

The fruit fly *Drosophila melanogaster* is an excellent model system for studying cell migration in the context of nervous system development since these processes are highly conserved and the fly is a versatile genetic tool. In addition, *Drosophila* embryo development as a whole is well characterized (Jennings 2011). Importantly, the *Drosophila* midline – an analogous structure to the vertebrate floor plate, which guides neuronal positioning and differentiation in an embryonic neural tube – is a coordinated system in which cells migrate and develop in distinctive patterns over time (Crews et al., 1992; Klämbt et al., 1991; Tessier-Lavigne et al., 1988). Thus, we can use the fly midline to gain insight into cell migration during mammalian development.

Midline development is a well-characterized process in *Drosophila*

Drosophila embryogenesis occurs over an approximately 17-hour period after egg laying (AEL), and is split into 17 morphologically distinct stages on the basis of the major developmental event occurring (neuronal differentiation or midline retraction, for instance). Midline cells are distinguished through expression of the transcription factor *single-minded* (*sim*), a required component for expression of midline-specific genes (Crews, 1998; Fontana & Crews, 2012; Wheeler et al., 2006). **Figure 1**, courtesy of Kearney et al., 2004, depicts the development of the *Drosophila* midline. Sim transcript is first found in the mesectodermal precursors of the midline around stage 5 (Crews, 1998). At stage 8-9 the midline is composed of an average of 16 cells per segment. Two hours later at stage 11, the first evidence of dorsal/ventral preference is observed when these cells cluster and form the midline primordium. By stage 13, these cells can be identified as midline glia (MG), neurons, and median neuroblasts (MNB). At this time the MG (all wrapper-positive) can further be defined as anterior MG (AMG) and posterior MG (PMG) on the basis of *runt* or *engrailed* (*en*) expression, respectively (Wheeler et al., 2006). AMG can also be identified via *wrapper* expression; thus, *wrapper* and *en* were used to divide segments into general anterior and posterior areas. Neurons - such as midline precursor neurons 1 and 3 (MP1, MP3), ventral unpaired median neurons (VUMs) and MNB - can be identified via differential gene expression (Wheeler et al., 2006). By stage 17, midline neurons and glia are matured and the embryo undergoes its next phase of life as a larva.

Specific and coordinated migration underlies the proper development of the midline, and a prominent example of midline cell migration is that of the MG. The MG must migrate and ensheathe the axons crossing the midline to form the MG-axon scaffold and commissures. Not only must the MG wrap around the axon as it is crossing the midline, but the correct MG must

ensheathe it; PMG undergo programmed cell death during embryonic development and play no role in axon ensheathment (Crews, 2010). Mutations in midline genes that alter pattern formation can lead to fused axon commissures, missing anterior/posterior commissures, or even the lack of a commissure entirely (Klämbt et al., 1991). The spatiotemporal location of cells is important for the proper structure and development of the midline, thus migration has to be a carefully coordinated and regulated event.

Drosophila Toll **plays a developmental role and shares homology to human Toll-like Receptors**

Nobel laureates Christiane Nüsslein-Volhard and Eric Wieschaus first characterized the *Toll* (*Tl*) gene in the late 1970s through their large-scale screen to identify genes involved in embryogenesis. To do so, they generated random mutations in flies via ethyl methanesulfonate (EMS) and examined mutant embryos and larvae for morphological defects in patterning. Larvae with a mutation in *Tl* were completely dorsalized. After these initial screens, continued characterization of the effect of *Tl* on development was conducted. It was found that the ventrally localized proteolytic processing of Spätzle (Spz), the Toll receptor ligand, could activate the Toll pathway. Toll activation occurs in a gradient and initiates a signaling cascade, eventually leading to Dorsal nuclear entry. Thus, Dorsal protein is distributed in a concentration gradient along the dorsal-ventral axis of the early embryo, setting expression limits of zygotic regulatory genes, which initiate the differentiation of various tissues (Anderson et al., 1985a; Anderson et al., 1985b, Morisalo & Anderson, 1995). Additionally, the Toll signaling pathway is involved in immunity, where the nuclear translocation of Dorsal-related immunity factor (Dif) results in the production of antimicrobial peptides (AMPs) (Valanne et al., 2011).

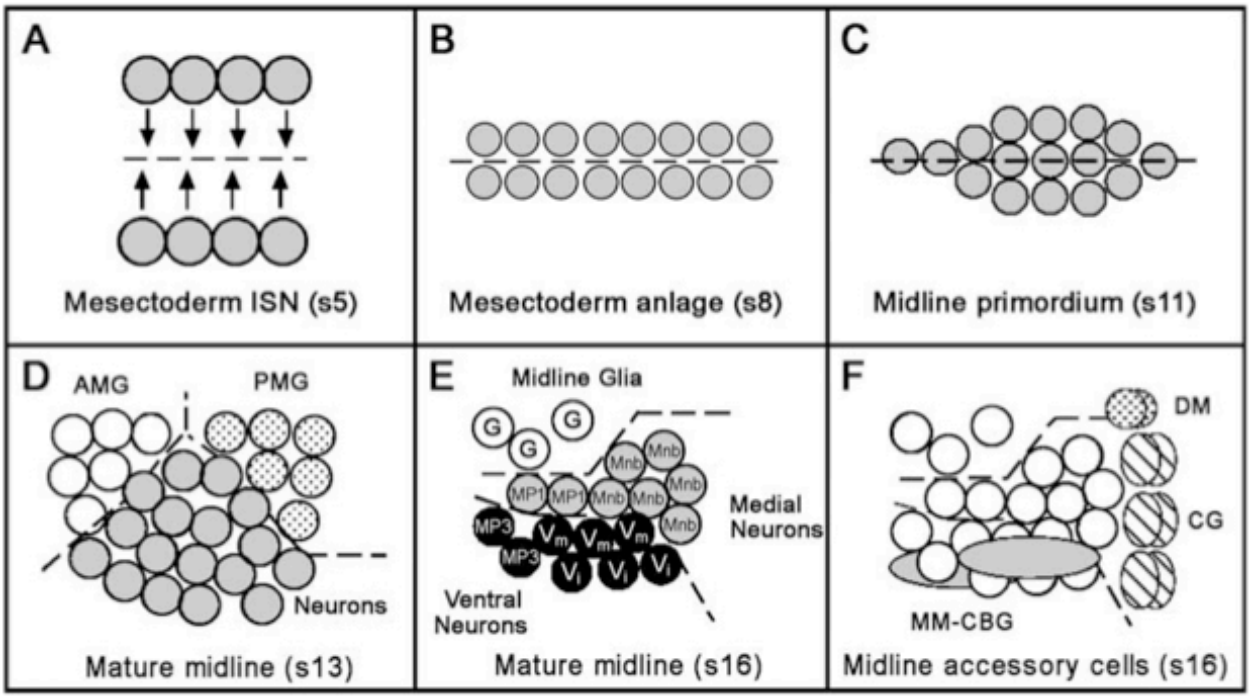

Figure 1: Schematic summary of CNS midline cell development. In all panels, a single segment is shown with anterior to the left. Embryonic stages are indicated by "s#". (A) Mesectoderm ISN stage, ventral view. Two stripes of mesectodermal cells reside on either side of the mesoderm in the blastoderm embryo (stage 5). Dotted line indicates ventral midline of embryo. There are four cells/segment on each side. Arrows represent how the mesectodermal cells move together at the ventral midline during gastrulation (stage 6) as the mesoderm invaginates. (B) Mesectoderm anlage stage, ventral view. During the mesectoderm anlage stage (stages 7–8), the mesectodermal cells meet at the midline and then undergo a synchronous cell division, resulting in 16 cells per segment. (C) Midline primordium stage, ventral view. During the midline primordium stage, midline cells rearrange from a two-cell wide planar array into a cell cluster. Midline cells within these clusters differ slightly in their dorsal/ventral positions. (D) Mature CNS midline cells, stage 13. Sagittal view, dorsal up. At stage 13, two populations of midline glial cells become evident. The anterior midline glia (AMG; open circles) are reduced by apoptosis but ultimately will ensheathe the commissures while all posterior midline glia (PMG; dotted circles) will undergo apoptosis. Midline neurons (shaded circles) occupy the space between and below the midline glia. Dotted lines separate the different cell groups. (E) Mature CNS midline cells, stage 16. Sagittal view, dorsal is up. The PMG have undergone apoptosis and are absent, whereas the AMG give rise to ~3 mature glia (G, open circles). Midline neurons have migrated to their final positions within the ganglion. Medial neurons include MP1 neurons (MP1, shaded circles) and the progeny of the MNB (Mnb, shaded circles). Ventral neurons include VUM motorneurons (Vm, black circles), VUM interneurons (Vi, black circles), and MP3 neurons (MP3, black circles). (F) Midline accessory cells shown in relation to midline neurons and glia (open circles). Two DM cells (dotted circles) lie atop the CNS near the midline channel, which is lined by six-channel glia (CG; hatched ovals). The two MM-CBG in each segment (shaded ovals) are closely associated with the ventral neurons.

The first indication that *Drosophila Toll* (*dToll*) shared a structural homolog in mammals came from work by Nobuo Nomura. In 1994, Nomura and colleagues were the first to clone and identify a Toll-like receptor (TLR) present in humans by comparing the sequence of their clone

to a database (Nomura et al., 1994). A later study in 1997 by Medzhitov and colleagues showed that a human TLR could activate the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) pathway, like *dToll*. This provided the first evidence that human *Toll* also played a role in immune function, indicating functional homology with *dToll* (Medzhitov et al., 1997). These works demonstrate that *dToll* is both structurally and functionally homologous to human *Toll*, suggesting a conserved evolutionary pathway.

Toll **is involved in immune function**

Mammalian TLRs are mainly known for immunological roles and act as pattern recognition receptors (PRRs) in innate immunity (Beutler & Rietschel, 2003; Iwasaki & Medzhitov, 2004; Takeda & Akira, 2005), specifically through activation of the NF-κB pathway. In *Drosophila,* a Gram-positive or fungal infection can trigger an immune response by activating the Toll pathway, in which cleaved Spätzle binds to the receptor, initiating an intracellular signal cascade which ultimately results in Dif and Dorsal activating transcription of drosomycin and other antimicrobials (Silverman et al., 2009). This pathway is similar to the mammalian interleukin-1 receptor pathway (IL-1R), which also induces a signaling cascade leading to the nuclear entry of NF-κB (homologue is Dorsal) and mediation of the immune response (Medzhitov, 2001).

Research within the past five years implicates a role for mammalian TLRs in tumor invasiveness and other central nervous system (CNS) pathologies (Carty & Bowie, 2011; Gooshe et al., 2014; Sarrazy et al., 2011; Trotta et al., 2014; Yoon et al., 2015). For instance, TLR4 is present in glioblastomas, and the activation of the TLR4 pathway led to increased cell proliferation. However, when activated with the Fas pathway - which escalates glioblastoma

pathogenicity and increases cell migration and proliferation when activated alone - cell migration and proliferation decreased (Sarrazy et al., 2011). The TLR4 pathway is able to modulate the proliferation and migration induced by the Fas pathway, as well as contribute to cell proliferation when activated. This suggests a role for TLR4 in regulating cell proliferation and migration, both in combination with other pathways and alone. Additionally, high TLR expression has been found in many cell types in the human CNS, and certain TLRs are proposed to play roles in CNS pathologies such as Alzheimer's disease, multiple sclerosis, and stroke (Carty & Bowie, 2011). Thus, we can use the fly to address gaps in knowledge concerning the roles of certain TLRs in a developmental context, and elucidate cellular migration mechanisms.

Toll belongs to a family of nine receptors in Drosophila

The *Drosophila* Toll family consists of nine closely related members: Toll, 18w, and Toll-3 through Toll-9. All contain a putative transmembrane domain as well as an extracellular domain composed of leucine-rich repeats (LRRs), which is C- or N-flanked by cysteine-rich motifs (**Figure 2**) (Bilak et al., 2003). LRRs help provide a structural framework for the formation of protein-protein interactions, while the flanking cysteine-rich regions stabilize the LRR (Kobe & Kajava, 2001). The N-terminus (start of the protein, amino-terminus) is extracellular, while the C-terminus (end of the protein, carboxyl-terminus) is intracellular. Toll-9 is the most structurally similar to mammalian TLRs in that it does not have an N-flanking cysteine-rich motif at the C-terminal end of the ectodomain (Bilak et al., 2003). Interestingly, Toll-9 is the only other family member besides Toll (and to a small extent, Toll-5) that can activate an immune response in *Drosophila*, shown in vitro (Ooi et al., 2002).

Studies indicate that the other Toll family members appear to play roles in development and other cellular processes. Most Toll family receptors are expressed at high levels during development in a variety of tissues undergoing cellular and morphogenetic movements and some are proposed to function as adhesion molecules (Eldon et al., 1994; Imler & Hoffmann, 2000; Kleve et al., 2006; Kolesnikov & Beckendorf, 2007). For instance, *Toll-8*, also known as *Tollo*, is necessary for the neural-specific induction of glycosylation in embryos, which may function in cellular communication during development (Seppo et al., 2003). *Toll-8* also acts in combination with *18w* and *Toll-6* to direct polarity and cell rearrangements during convergent extension (Paré et al., 2014). In addition, *Toll-6* and *Toll-7* are expressed in the CNS and are required for locomotion, neuronal survival, and motor axon targeting (McIlroy et al., 2013). This suggests that the immunity role of Toll family members is the exception rather than the rule, since the majority of receptors are needed for proper embryonic development.

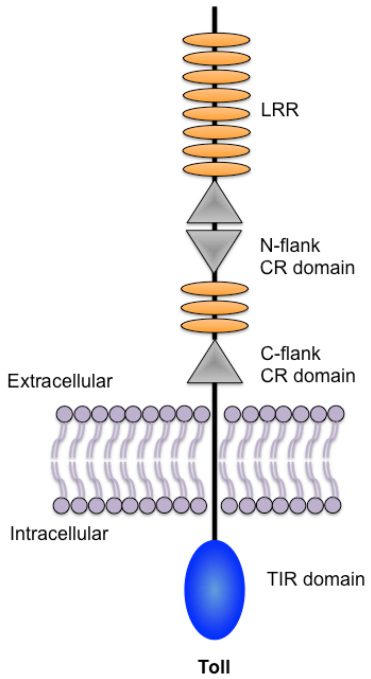

Figure 2: Structure of *Drosophila* Toll. The ectodomain is comprised of LRRs flanked by cysteine-rich (CR) domains. The intracellular Toll/IL-1R (TIR) domain interacts with adaptor proteins after activation and signals through NF-κB-like molecules. The Toll family of receptors follows this structural pattern with the exception of Toll-9, which does not have an N-flanking CR domain at the C-terminal end of the ectoderm.

***18-wheeler (18w)* contributes to embryonic development and cellular movement**

18w (also known as *Toll-2*) is present in early embryonic development in a pattern reminiscent of segment polarity genes such as *engrailed,* and is frequently expressed in regions undergoing cell movement (Eldon et al., 1994). Most homozygous mutants die as larvae, although surviving adults typically display antenna, wing, and leg deformities. Since *18w* is expressed in many tissues undergoing movement, it is probable that defective cell movements cause these appendage anomalies. Therefore, larvae likely die due to an accumulation of these morphogenetic movement defects (Eldon et al., 1994). As stated, the expression pattern of *18w* resembles that of other segment polarity genes such as *en* and transcription factors Eve and Runt; these polarity genes are critical for the correct anteroposterior patterning of embryonic segments, and mutants for these genes show deletions in segment structures or other abnormalities (Patel et al., 1989). *18w* may be playing a similar role.

Regarding cellular movement, *18w* has also been shown to regulate apical constriction of the salivary gland through the Rho-GTPase pathway, which is involved in many aspects of cytoskeletal dynamics and cell movement (Van Aelst & D'Souza-Schorey, 1997). Loss of function *18w* mutants showed defects in the synchronicity of salivary gland invagination and positioning due to a failure of cells to migrate properly. Conversely, *18w* overexpression causes an upregulation in Rho signaling, as observed by the two-fold increase in a marker of Rho signaling activation, phospho-Spaghetti squash (P-Sqh) (Kolesnikov & Beckendorf, 2007). *18w* also plays a role in the migration of follicle cells in the *Drosophila* ovary, as females carrying an *18w* loss of function (LOF) mutant clone in their ovaries show delayed follicle cell migration. Furthermore, their eggs laid had structural and morphological defects; some eggs were even deflated (Kleve et al., 2006).

Additionally, *18w* in combination with *Toll-6* and *Toll-8* regulates convergent extension in embryo development. Analysis of single, double, and triple mutants for these genes revealed that the more Toll receptors lost, the greater the occurrence of defects in tissue elongation, cell intercalation, and edge formation. However, embryos lacking *18w* alone showed defects in edge formation, revealing that neither *Toll-6* nor *Toll-8* can be fully redundant in this process (Paré et al., 2014). *18w* may also have a role as a cell adhesion molecule, as it promotes cell aggregation *in* vitro when expressed in Schneider 2 (S2) cells, which in their un-altered state are non-adhesive (Eldon et al., 1994). This proposed role as an adhesion molecule however, needs more experimental evidence to confirm.

Many examples are given indicating the involvement of 18w in cellular movement; however, it is unknown if this receptor is involved in cellular migration in the midline. The midline is a highly choreographed and well-documented developmental system, and perturbing 18w function could reveal novel roles for the receptor in midline development. Further explorations of the mechanisms linking *18w* signaling to dynamic cellular migration changes will provide insight into this highly conserved developmental process and may provide more understanding of neurodevelopmental diseases involving aberrant cellular migration.

METHODS AND MATERIALS

Drosophila strains and genetics

Drosophila strains used included: *18w (Δ7-35)* (Eldon et al., 1994)*, CyO-ftz-lacZ* (Roark et al., 1995), *sim-Gal4* (Xiao et al., 1996), and *UAS-tau-GFP* (Wheeler et al., 2006).

Generation of fly strains

18w (Δ7-35) was balanced with *CyO*, to generate *18w (Δ7-35)/CyO*. These flies were crossed to *STG* on the third chromosome to generate *18w (Δ7-35)/CyO; STG/STG* [*18w (Δ7-35)/CyO; STG*]. *18w (Δ7-35)* was also balanced with *CyO-ftz-lacZ*, to generate *18w (Δ7-35)/CyO-ftz-lacZ*. These flies were crossed to *STG* on the third chromosome to generate *18w (Δ7-35)/CyO-ftz-lacZ; STG/STG* [*18w (Δ7-35)/CyO-ftz-lacZ; STG*].

Sources of cloned DNA for in situ hybridization

cDNA clones from the *Drosophila* Gene Collection (Open Biosystems, versions 1.0 and 2.0) (Stapleton et al., 2002) were used to prepare digoxigenin-labeled or biotin-labeled *in situ* hybridization probes for *en, gsb, odd, wrapper, wor,* and *zfh1*. *18w* was PCR-amplified from genomic DNA using a previously published gene-specific primer set incorporating the T7 promoter sequence (5′-TGCAACTGCTCAATCTCACC-3′ and 5′-taatacgactcactatagggagaTACTCCGACTCGATGCTGTG-3′) (Paré et al., 2014).

In situ hybridization, immunostaining, and microscopy

Embryo collections, in situ hybridization, immunostaining, and confocal imaging were performed as previously described (Kearney et al., 2004; Wheeler et al., 2006; Wheeler et al, 2008). Primary antibodies used for staining were rabbit anti-β-galactosidase (Invitrogen) and mouse anti-Tau (Sigma). Alexa Fluor-conjugated secondary antibodies (Molecular Probes) were used as secondaries. The Tyramide Signal Amplification System (Perkin-Elmer) was also used

for some experiments depending on fluorescent signal strength. Flies ranging from stages 11 through 16 of development were used for most experiments, and were methanol/formaldehyde-fixed prior to use. All embryos were collected on grape juice plates streaked with fresh yeast paste and kept at -20°C until ready to use.

RESULTS

Generation of antisense RNA probes for *18w*

In order to visualize *18w* expression patterns *in vivo*, an antisense RNA probe was transcribed from genomic DNA (gDNA) using a gene-specific set of primers, which included the T7 promoter sequence at the 5' end of the reverse primer, and T7 RNA polymerase. Typically, this probe would be prepared from the cDNA clone located in the *Drosophila* Genome Collection; however, this method was unsuccessful in that a labeled, antisense probe that recapitulated the known expression patterns of *18w* could not be obtained. Thus, gDNA was isolated from both w^{1118} and *sim-Gal4; UAS-τ-GFP* (*STG*) flies, both of which have a wild-type copy of *18w*. Once the gDNA was isolated, *18w* was PCR-amplified with primers that incorporated the T7 promoter sequence into the PCR product. After purifying the PCR product, an in vitro transcription reaction was run using T7 RNA polymerase, which transcribes from the 5' to 3' direction. The polymerase can also incorporate certain labels into the product, such as DIG or biotin in this specific case, which can be used for later detection experiments. This reaction produces a labeled antisense RNA probe that can bind mRNA (**Figure 3A**).

After the PCR amplification process, a small amount of product was run out on a 1% agarose gel to confirm that only one band was seen at approximately 3300 base pairs. Indeed, only one band was seen at the estimated size for both the w^{1118} and *STG* samples, suggesting that *18w* had been specifically amplified (**Figure 3B**). A higher concentration of PCR product was obtained in the *STG* sample, therefore it was used for in vitro transcription and labeling. After the transcription procedure both the biotin- and DIG-labeled probes were run out on a gel to ensure that RNA was made (**Figure 3C**).

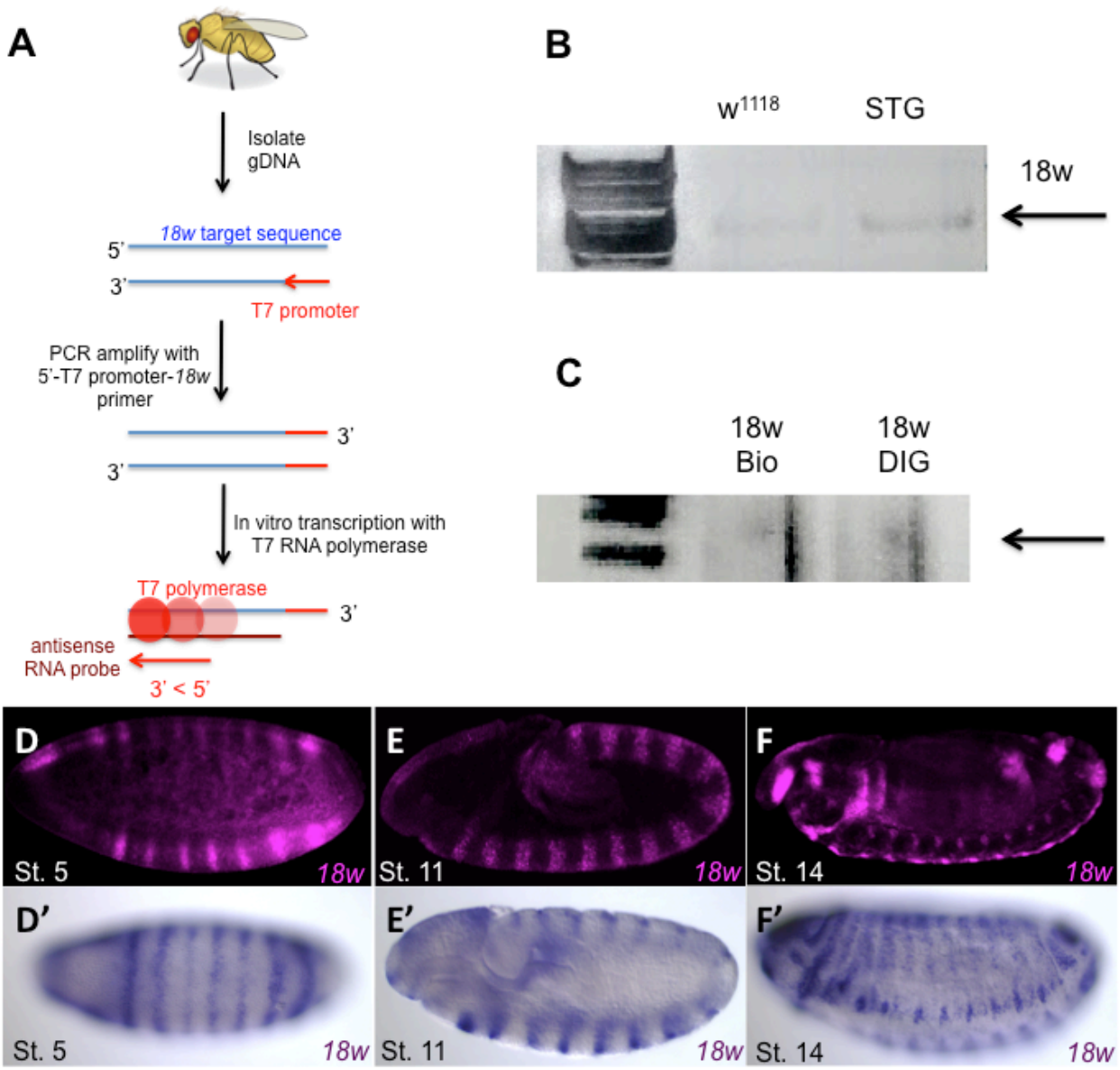

Figure 3: Generated *18w* RNA probes recapitulate the expression patterns seen in the BDGP in situ collection. (A) Schematic of *18w* RNA probe generation through gDNA PCR amplification and in vitro transcription. (B) *18w* was PCR amplified from w^{1118} and STG fly gDNA and the product was run out on a 1% agarose gel. Band sizes are approximately the appropriate length. (C) Generated probes were run out on a gel. RNA is present at the expected length for both DIG and biotin probes. (D-F) *In situ* hybridization of STG embryos at different developmental stages reveals similar expression patterns of *18w* to established experiments. The *18w* DIG probe was used and conjugated to Cy3 for visualization. Embryos are anterior forward, sagittal view. (D'-F') BDGP alkaline phosphatase *in situ* experiments used as the reference for *18w* expression. Embryos are anterior forward.

Embryos from early (stage 5) to late (stage 14) developmental stages were hybridized with biotin- and DIG-labeled probes specific to *18w*, and compared to known *18w* embryonic expression patterns from the Berkeley *Drosophila* Genome Project (BDGP) *in situ* database (Hammonds et al., 2013; Tomancak et al., 2002; Tomancak et al., 2007). Embryos from all developmental stages collected (not all stages shown, **Figure 3D-F**) exhibited similar expression patterns to similarly aged BDGP embryos (**Figure 3D'-F'**). Thus, the *18w* probes transcribed can be used to successfully target *18w* mRNA and characterize its expression in subsequent experiments.

18w is present in the posterior of segments and expressed in median neuroblasts

Prior research conducted in the lab includes a transcriptome analysis of 3.7*sim-Gal4; UAS-mCD8.GFP* flies using fluorescence activated cell sorting (FACS) to isolate midline cells at two different developmental periods (6-8 hours AEL and 14-16 hours AEL). The earlier stage represents the period when midline neurons are beginning to differentiate, while at the later stage, neurons and glia are mature and well-differentiated. The fly strain used the enhancer of *sim* to drive the expression of GFP specifically in the midline. The two midline cells samples were then processed for RNA sequencing. Data generated from these experiments were analyzed to identify distinctive traits of midline cell types, as well as to examine how these cells acquire their differentiated state (Fontana & Crews, 2012). Unpublished data from that study revealed elevated levels of *18w,* as well as other Toll family receptors, in the midline at these two developmental time points (**Table 1**), corroborating their known roles in embryonic development.

Toll Family Receptor	FPKM 6-8 hr AEL	FPKM 14-16 hr AEL
Toll (Tl)	396.295	81.2687
18w	107.828	5.02417
Toll-6	55.1867	1.75523
Toll-7	144.348	93.6888
Tollo (Toll-8)	55.9107	74.1280

Table 1: FACS purified midline cells at 6-8hr (stages 11/12) and 14-16hr (stage 16) AEL were purified and RNA-sequenced. Expression levels are indicated by FPKM (fragments per kilobase of exon per million fragments mapped); a cutoff of 5.0 was used for transcript detection.

The midline is composed of multiple cell types, including anterior/posterior midline glia, VUMs, and the MNB. In order to determine the precise spatial and temporal expression of *18w* and the other Toll family members within the midline, embryos were stained using a combination of antibodies and antisense RNA probes for specific midline cell types (**Table 2**). Anterior and posterior midline cells were labeled using *wrapper* and *en*, respectively. In staining for *18w* and *Tl*, I found that *18w* is present in the posterior part of a segment, while *Tl* is present in the anterior of a segment, with little overlap between the two (**Figure 4, and data not shown**). These distinct expression patterns suggest a possible role for *Tl* and *18w* in compartmentalization within the midline. As both are shown to play roles in cellular movement in other cell types *in vivo*, I propose that *Tl* and *18w* may act to guide migrating midline cells to their final destinations within a segment (Paré et al., 2014; Wang et al., 2005).

During mid-development (stages 11-12), *18w* is expressed in the posterior portion of each segment. There is little to no co-localization with *wrapper* (midline glia), *gsb* (H-cell, H-cell sib, MP3), *odd* (MP1), or *zfh1* (motor VUMs) (**Figure 5A-C, E**). However, *18w* expression overlaps with *wor*, a marker of MNB (**Figure 5D**). During late development (stages 14-16), *18w* expression is again observed in the posterior of segments and does not co-localize with *wrapper, gsb, odd,* or *zfh1,* but continues to overlap with *wor* expression in the MNB (**Figure 5F-J**). Thus, it appears that *18w* is expressed specifically in the MNB.

Midline cell gene expression			
All midline cells	*sim*	MNB	*en*, *wor*
AMG	*runt, wrapper*	MP1; MP1 neurons	*lim3,odd*
H-cell	*gsb*, *ple*, *tup*	MP3	*gsb*
H-cell sib	*CG13565*, *gsb*, *VGlut*	MP4-6	*en*
iVUMs	*en*, *gad1*	PMG	*en*, *mas, wrapper*

Table 2: Midline cells can be identified by their expression of certain genes using either RNA probes or antibodies.

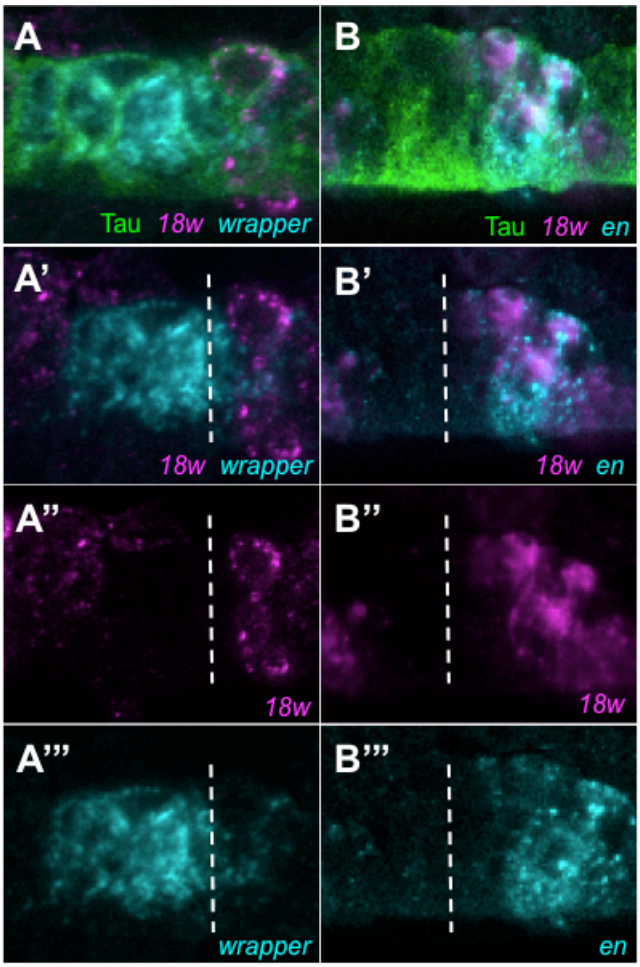

Figure 4: *18w* is expressed in the posterior of segments in stage 11 embryos. (A-A''') FISH of *STG* embryos using *18w* (magenta) and *wrapper* (cyan) probes along with anti-tau (green) to visualize the midline. The same segment is depicted without midline staining (A') to show *18w* posterior and *wrapper* anterior expression more clearly. The segment is positioned anterior to the left, with the dashed white line serving as an approximate divider between anterior and posterior areas. (B-B''') FISH of *STG* embryos using *18w* (magenta) and *en* (cyan) probes along with anti-tau (green) to visualize the midline. The same segment is depicted without midline staining (B') to show the overlap of *18w* and *en* expression in the posterior. The segment is positioned anterior to the left, with the dashed white line serving as an approximate divider between anterior and posterior areas.

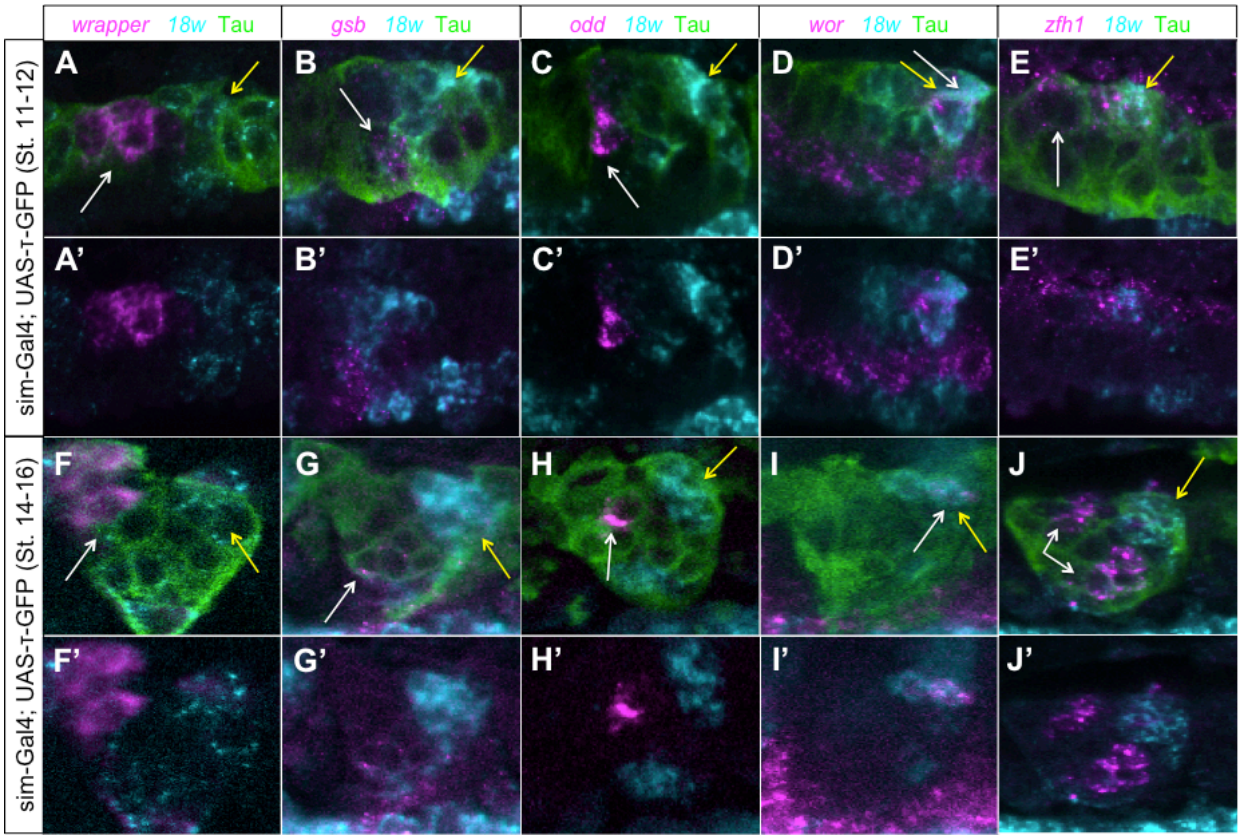

Figure 5: *18w* is expressed in the MNB during mid- to late-development. (A-E) FISH of *STG* stage 11-12 embryos using *18w* (cyan) and other cell-specific marker (magenta) probes. Anti-tau (green) was used to visualize the midline. *18w* and *wor* (D) appear to overlap. Segments are positioned anterior to the left at 40x magnification. White arrows point to cell-specific marker expression, and yellow arrows point to *18w* expression. (A'-E') The same segments from A-E are shown without midline staining to emphasize probe staining and localization. (F-J) FISH of *STG* stage 14-16 embryos using the same probes and color scheme as in A-E, with anti-tau (green) utilized to visualize the midline. Segments are positioned anterior to the left at 40x magnification. White arrows point to cell-specific marker expression, while yellow arrows point to *18w* expression. (F'-J') The same segments from F-J are shown without midline staining to emphasize probe staining and localization.

Overexpression of *18w* in the midline causes defects in midline structure

Given the distinct posterior location of *18w* in midline segments and the work of others which suggests *18w* has roles in cell migration, I hypothesized that *18w* may have a role in guiding midline cells to their correct locations within a segment during development. As such, a loss or gain of 18w expression may disrupt cell localization and migration in distinct ways. For example, overexpression of *18w* in the midline may cause anterior cells to incorrectly migrate towards the posterior of a given segment, while loss of *18w* may cause posterior cells to migrate anteriorally within a segment due to loss of positioning cues from *18w* (**Figure 6**).

In order to test this hypothesis, *18w* was overexpressed in the midline using *UAS-18w* crossed to fly background *STG*, which drives the expression of GFP in all midline (*sim*-positive) cells. Using this system, *18w* is expressed in every cell that expresses *sim*, namely, all midline cells. *en* was used as a posterior marker to assess cell positioning. *UAS-18w* was also crossed to w^{1118}, which has the same genetic background as the *STG* strain, to ensure that any differences seen in midline development were due to the presence of *18w* being overexpressed by *sim-Gal4*, and not to the genetic background of *STG*. Additionally, embryos from the cross *sim-Gal4; UAS-tau-GFP* x w^{1118}, served as the controls for the experiments described within this section. As another control, *UAS-18w* embryos were stained for *18w*, *en,* and other markers listed in **Table 2**, and no appreciable difference was noted between these embryos and *STG* embryos (data not shown).

Compared to control embryos (*sim-Gal4/+; UAS-tau-GFP/+*) at comparable developmental stages (**Figure 7A-B''**), embryos overexpressing *18w* (*sim-Gal4/UAS-18w; UAS-tau-GFP/+*) showed several notable differences (**Figure 7C-F''**). In these mutant embryos, all *sim* expressing cells are expected to express tau-GFP, as well as overexpress *18w*. However, tau

expression appeared lower in some midline cells (**Figure 7C', 7D'**, **7E'**, yellow arrows). Further, some segments expressed a "hole" in *18w* overexpression that overlapped with the lowered tau expression (**Figure 7C'', 7D''**, **7E''**, pink arrows). In addition to the curious hole in *18w* expression, some segments also revealed fusion defects (**Figure 7E', 7F'-F'')**.

In several segments, some cells appeared to be fused together (**Figure 7E'**, yellow arrow). In extreme cases, segments appeared to be fused together (**Figure 7F'-F''**, white arrows).

Examining overall midline morphology, there appear to be differences between control and overexpression embryos when relatively stage-matched. Around stage 11, overall midline cells in the *sim-Gal4/UAS-18w*; *UAS-tau-GFP/+* embryos appear to be disorganized and elongated compared to controls (compare **Figure 7A-A''** to **7C-D''**). Around stage 14, overall midline structure is disorganized, including improper fusion events between cells within and between segments (compare **Figure 7B-B''** to **7E-6F''**). The data presented show that overexpressing *18w* affects midline morphology during development.

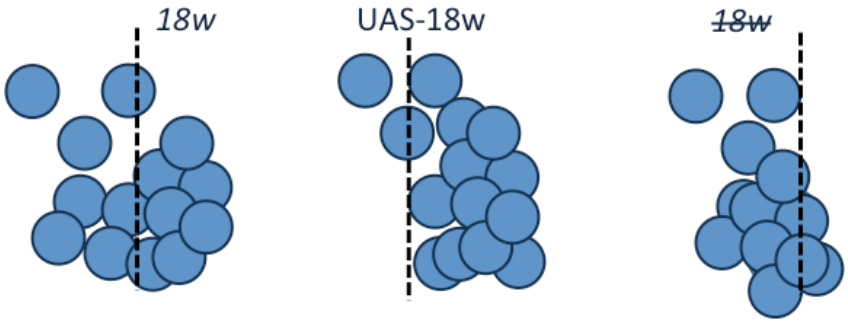

Figure 6: Schematic depicting hypothesized normal expression of *18w* (left), overexpression of *18w* (center) and mutation or deletion of *18w* (right). The circles represent cells, and the dotted line represents an approximate boundary between the anterior and posterior of a segment.

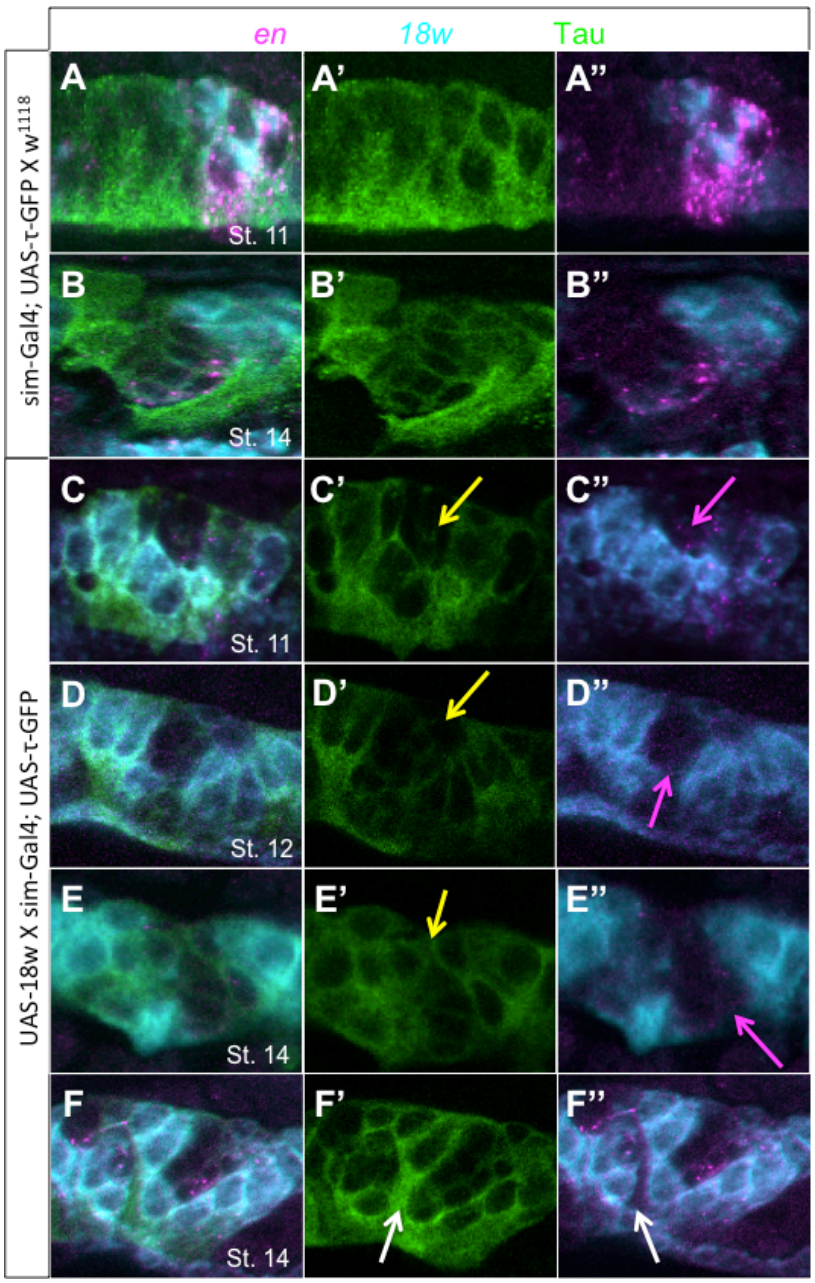

Figure 7: Overexpressing *18w* leads to overall midline structure abnormalities relative to controls. (A-B'') FISH of *STG* x w^{1118} stage 11 and 14 control embryos using *18w* (cyan) and *en* (magenta) probes at 40x magnification. Anti-tau (green) was used to visualize the midline. These probes and color scheme is used for the rest of the panels in this figure. A' and B' reveal midline structure only, while A'' and B'' show *en* and *18w* expression. (C-C'') FISH of *UAS-18w* x *STG* at stage 11. The yellow arrow in C' points to a hole in tau expression, and the pink arrow in C'' points to a hole in *18w* expression. Midline morphology appears disorganized. (D-D'') FISH of *UAS-18w* x *STG* at stage 12. The yellow arrow in D' points to an area with less tau expression, that overlaps with the area of no *18w* expression seen in D'' (pink arrow). Midline morphology appears disorganized. (E-E'') FISH of *UAS-18w* x *STG* at stage 14. The yellow arrow in E' points to both a hole in tau expression and to what appears to be an improper cell fusing event. In E'' the pink arrow points to an area with no *18w* expression. Overall midline structure looks abnormal. (F-F'') FISH of *UAS-18w* x *STG* at stage 14. The white arrow in F' points to an apparent fusion between two segments. The white arrow in F'' indicates the same event. Once again, overall midline structure appears disorganized compared to the stage-matched control.

Heterozygous mutants of *18w* do not appear to cause overt defects in midline structure

Since overexpression of *18w* in the midline revealed structural abnormalities, I then investigated if complete/partial loss of function mutations of *18w* would lead to defects in midline development. These experiments used the *18w (Δ7-35)* line. The Δ7-35 mutant carries a large deletion of about 2.2kb, which removes about 1.7 kb of the open reading frame, resulting in a severe lack of function allele (Eldon et al., 1994). However, *18w* transcript is still produced in these mutants. As a homozygous or trans-heterozygous mutation, it is sub-lethal. Death of these mutants occurs during larval development, with approximately 0.5% of mutant embryos surviving into early adulthood (Eldon et al., 1994).

Crosses within the *18w (Δ7-35)/CyO; STG* strain produced embryos with the following genotypes and frequencies, based on Mendelian inheritance: *18w (Δ7-35)/18w (Δ7-35); STG* (25%), and *18w (Δ7-35)/CyO; STG* (50%), and *CyO/CyO; STG* (25%). The first genotype is a homozygous mutant, while the second is heterozygous and for these experiments served as the control genotype. The last genotype is lethal, since balancer chromosomes contain a lethal recessive allele. Thus, we would expect to see approximately 66% *18w (Δ7-35)/CyO; STG* (heterozygous) embryos and 33% *18w (Δ7-35)/18w (Δ7-35); STG* (mutant) embryos. In the approximately 300 embryos examined from crosses within this strain, none were convincingly negative for *18w* staining (data not shown). This is expected, as previous work indicates that homozygous *18w (Δ7-35)/18w (Δ7-35)* embryos produced a truncated transcript (Eldon et al, 1994). Alternatively, the *18w (Δ7-35)/18w (Δ7-35)* genotype could be lethal on our *STG* background. Further, there was no appreciable difference between wild-type *STG* and any of the embryos from the *18w (Δ7-35)/CyO; STG* midline structures when stained with posterior marker *en*, glial marker *wrapper*, and anti-tau (**Figure 8A-D**).

The lack of a midline phenotype in any of the embryos from the *18w (Δ7-35)/CyO; STG* strain suggests that loss of *18w* does not visibly alter midline morphology, or that *18w (Δ7-35)/18w (Δ7-35); STG* is lethal. It is imperative to differentiate between these possibilities, which requires differentiation between *18w (Δ7-35)/CyO; STG* (heterozygous) embryos and *18w (Δ7-35)/18w (Δ7-35); STG* (homozygous mutant) embryos. In order to distinguish heterozygous and homozygous mutants, the *18w (Δ7-35)* line used was crossed to marked balancer *CyO-ftz-lacZ*, as well as to *STG* on the third chromosome. Ignoring the *CyO-ftz-lacZ/CyO-ftz-lacZ* genotype which is lethal, the two possible genotypes on the second chromosome are *18w (Δ7-35)/18w (Δ7-35)* or *18w (Δ7-35)/CyO-ftz-lacZ.* A homozygous mutant will not stain for β-gal, whereas a heterozygous mutant will due to the marked balancer. Approximately 300 embryos were stained and analyzed for β-gal expression, and all were positive for β-gal. This means that all the embryos examined were heterozygous mutants for *18w (Δ7-35)* (**Figure 8E-F**). Once again, no noticeable differences were observed in midline structure. It is possible that if more embryos were analyzed, a few homozygous mutants would be observed. However, these data suggest that *18w (Δ7-35)/18w (Δ7-35); STG* is lethal, and that a single wild-type copy of *18w* is sufficient for processes regulating normal midline development and organization.

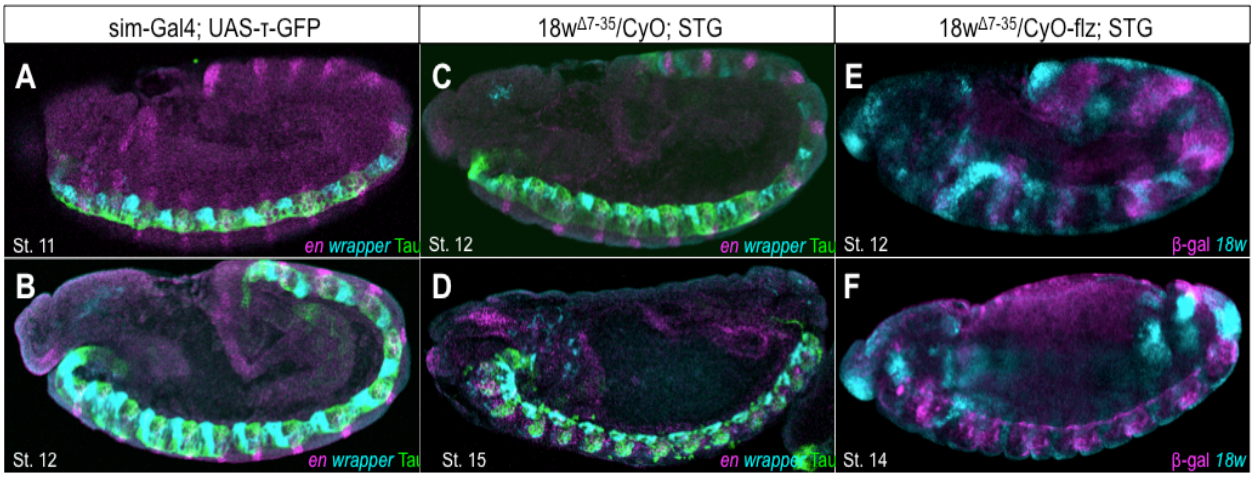

Figure 8: *18w* heterozygous mutants do not appear to affect midline morphology. All embryos shown are anterior forward, sagittal view at 10x magnification. (A-B) FISH of *STG* control embryos at stage 11 and 12 using *en* (magenta), *wrapper* (cyan), and anti-Tau to visualize the midline (green). *en* stains the posterior of embryo segments while *wrapper* stains the midline glia. At this point in development glial staining is more prominent in the anterior of segments. (C-D) FISH of *18w (Δ7-35)/CyO; STG* embryos at stages 12 and 15 using *en* (magenta), *wrapper* (cyan), and anti-Tau to visualize the midline (green). There are no overt differences in midline morphology relative to the control. (E-F) FISH of *18w (Δ7-35)/CyO-ftz-lacZ; STG* embryos at stage 12 and 14 using β-gal to detect lacZ (magenta) and *18w* (cyan). All embryos examined that had β-gal staining also stained positive for *18w*, indicating that these embryos are heterozygous mutants. No differences in midline morphology compared to controls were noticed.

DISCUSSION

The data presented in this chapter corroborate RNA-sequencing results (Fontana & Crews, 2012) by demonstrating that *18w* is present in the midline during the developmental stages previously analyzed. I show that *18w* is localized specifically to the posterior of embryonic segments in the midline and is also present in the MNB. Midline neurons are generated in part by the asymmetric stem cell divisions of the MNB, which gives rise to five to eight GABAergic neurons during embryogenesis (Bossing & Technau, 1994; Truman et al., 2004). During development, these cells do not simply differentiate and remain where they are; they must migrate to their final destination. Cell migration depends on cells recognizing both their position in the environment and their polarity, which may involve detecting gradients of signaling molecules directing the cell where and how to move. It is possible that *18w*, specifically confined to the posterior of segments, provides guidance cues to MNB progeny to localize the newborn cells properly within the developing segment.

The expression pattern of *18w* resembles that of other segment polarity genes (Patel et al., 1989), and *18w* is shown to play roles in cell migration in multiple fly cell types, such as the follicle cell epithelium (Eldon et al., 1994; Kleve et al., 2006; Kolesnikov & Beckendorf, 2007). My data support a model where 18w plays a role in cell migration during embryogenesis, as overexpressing *18w* specifically in the midline causes abnormalities in midline structure and cell localization, as also shown in this chapter. Furthermore, *Tl* and other Toll family receptors, including *18w*, are proposed to act as adhesion molecules (Eldon et al., 1994; Imler & Hoffmann, 2000; Kleve et al., 2006; Kolesnikov & Beckendorf, 2007). Thus, it is possible that *18w* acts as an adhesion molecule to support migration in the midline.

When *18w* is overexpressed in the midline using *sim-Gal4,* some cells and segments appeared to be improperly fused together, and the overall structure of the midline seemed disorganized. It is possible that the ubiquitous expression of *18w* in the midline disrupts proper cell migration, since *18w* cues are no longer restricted to the posterior regions of midline segments. As a result, the cells may not differentiate appropriately, instead differentiating into another cell type based on signals from its incorrect environment. Another interesting phenotype of the *18w* overexpression embryos are the regions of reduced tau expression (both tau and 18w are transgenes driven by *sim-Gal4* for overexpression in the midline). One possibility is that the disrupted midline cell development contributes to the low presence of tau expression from the *UAS-tau-GFP* transgene. It is also possible that *STG* is not fully driving the expression of *18w*, and thus also not driving expression of *tau-GFP* in this experiment, which would explain the holes in expression seen for both *tau-GFP* and *18w*.

These data lend support to my hypothesis that *18w* is acting as an adhesion molecule promoting migration in the midline. It would be interesting to repeat these experiments using different cell-specific markers to determine if *18w* is affecting the positioning of specific midline cells. If *18w* is promoting migration and it is overexpressed, cell mislocalization, perhaps towards the posterior of segments, is probable. Thus, typically anterior cells such as the AMG might end up positioned more posterior than expected. This would contribute to the general disorganized midline structure seen. Although more experiments are necessary to determine the mechanisms behind the midline disorganization and fusion events seen, the overexpression experiments provide the insight that *18w* does have an effect on midline development.

A homozygous mutant for *18w* (*18w (Δ7-35)/18w (Δ7-35); STG/STG*) was not observed in the loss of function (LOF) experiments performed, and thus I cannot determine if loss of *18w*

in the midline alters midline development. There are several other LOF mutations of *18w* available that can be used in combination with Δ7-35 to produce a LOF, potentially less lethal mutant. Although in experiments like this, pleiotropy is a concern. Any observed defects in the midline development of mutants where *18w* is deleted throughout the entire embryo may be non-cell autonomous. That is, defects in other cell types may have an indirect effect on the midline. Thus, it would take further experimentation to determine if any observed effects are directly caused by loss of *18w* specifically in the midline.

In order to evaluate the effect of *18w* deletion on the midline, it is necessary to either generate a homozygous or transheterozygous mutant, which would assure *18w* loss of function, or use the power of clustered regularly interspaced short palindromic repeats (CRISPR) to generate a mutant. A recent advancement in genome editing, CRISPR could be implemented to create an *18w* mutant by targeting its midline enhancer region, potentially disrupting *18w* only in the midline. This would remove any concerns about pleiotropy, and allow for the assessment of the effect of *18w* LOF on midline development and cellular migration. In addition, CRISPR can be used to tag *18w* with a fluorescent marker *in vivo*, which would allow for live imaging of the protein. This could provide exciting insight into the roles *18w* has in midline cell migration, differentiation, and development.

It is possible that there were subtle defects in midline development of the heterozygotes (*18w (Δ7-35)/CyO; STG/STG*), however no appreciable differences were noticed between controls and these heterozygous mutants. Thus, it seems that one wild-type copy of *18w* is sufficient to maintain proper midline organization and development through the stages analyzed.

The experiments presented in this chapter provide evidence that *18w* is involved in midline development and future experiments were proposed that could provide knowledge

concerning the effect of *18w* loss on the midline. Further analysis is required to determine how *18w* is acting to guide cell migration (is it acting as an adhesion molecule?), why it is expressed in the MNB, and how a loss of function mutant would affect midline structure. Elucidating the role *18w* has to play in fly development may translate into knowledge about how the TLRs in humans contribute to both development and neurological disease, and may suggest new targets for therapy generation. Since the Toll family of receptors in flies is as a whole homologous to human TLRs, discovering new pathways and functions for these receptors in flies could shed light on novel roles and pathway involvement, or importantly, reveal druggable targets in humans for different disorders. For instance, since TLR4 can regulate proliferation and migration in glioblastoma induced by the Fas pathway as well as contributing to increased cell proliferation when activated, it would be insightful to study the mechanisms behind this further. Does TLR4 activation go through a pathway that inhibits the Fas pathway? Does it produce a byproduct of activation that inhibits the Fas pathway? Answering these questions may lead to the discovery of novel druggable targets to reduce the pathogenicity and metastasis of glioblastoma. This is only the beginning. Since high TLR expression has been found in many cell types in the human CNS, gaining insight into how the receptors function outside of immunity would be highly informative and perhaps contribute to our understanding of neurological disease. The fly can be used to address these gaps in knowledge and contribute to our understanding of the multiple roles of this large family of receptors.

REFERENCES

Anderson, K. V., Jürgens, G., & Nüsslein-Volhard, C. (1985a). Establishment of dorsal-ventral polarity in the Drosophila embryo: Genetic studies on the role of the Toll gene product. *Cell* **42**, 779–789.

Anderson, K. V., Bokla, L., & Nüsslein-Volhard, C. (1985b). Establishment of Dorsal-Ventral Polarity in the Drosophila Embryo : The Induction of Polarity by the Toll Gene Product. *Cell* **42**, 791–798.

Azevedo, F. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., Jacob, F. W., Lent, R., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neuro*. **513**, 532–541.

Beutler, B. & Rietschel, E. T. (2003). Innate immune sensing and its roots: the story of endotoxin. *Nat*. *Rev*. *Immunol*, **3**, 169–176.

Bilak, H., Tauszig-Delamasure, S., & Imler, J. L. (2003). Toll and Toll-like receptors in Drosophila. *Biochem*. *Soc*. *T*. **31,** 648–651.

Bossing, T. & Technau, G. M. (1994). The fate of the CNS midline progenitors in Drosophila as revealed by a new method for single cell labelling. *Development* **120**, 1895–1906.

Carty, M. & Bowie, A. G. (2011). Evaluating the role of Toll-like receptors in diseases of the central nervous system. *Biochem. Pharmacol.* **81**, 825–837.

Casanova, M. F. (2014). Autism as a sequence: from heterochronic germinal cell divisions to abnormalities of cell migration and cortical dysplasias. *Med. Hypotheses* **83**, 32–38.

Crews, S. T. (1998). Control of cell lineage-specific development and transcription by bHLH-PAS proteins. *Genes Dev*. **12**, 607–620.

Crews, S. T. (2010). Axon-glial interactions at the drosophila CNS midline. *Cell Adhes*. *Mig*. **4**, 67–71.

Crews, S., Franks, R., Hu, S., Matthews, B., & Nambu ,J. (1992). Drosophila single-minded gene and the molecular genetics of CNS midline development. *J. Exp. Zool.* **261**, 234–244.

Eldon, E., Kooyer, S., Evelyn, D. D., Duman, M., Lawinger, P., Botas, J., & Bellen, H. J. (1994). The Drosophila 18 wheeler is required for morphogenesis and has striking similarities to Toll. *Cell* **899**, 885–899.

Fontana, J. R. & Crews, S. T. (2012). Transcriptome analysis of Drosophila CNS midline cells reveals diverse peptidergic properties and a role for castor in neuronal differentiation. *Dev. Biol.* **372**, 131–142.

Gooshe, M., Abdolghaffari, A., Gambuzza, M., & Rezaei, N. (2014). The role of Toll-like receptors in multiple sclerosis and possible targeting for therapeutic purposes. *Rev. Neurosci.* **25**, 713-739.

Hammonds, A. S., Bristow, C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E., & Celniker, S. E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. *Genome Biol*. **14**, R140.

Imler, J. & Hoffmann, J. A. (2000). Signaling mechanisms in the antimicrobial host defense of Drosophila. *Curr. Opin. Microbiol.* **3**, 23–28.

Iwasaki, A. & Medzhitov, R. (2004). Toll-like receptor control of the adaptive immune responses. *Nat*. *Immunol*. **5**, 987–995.

Jennings, B. H. (2011). Drosophila – a versatile model in biology & medicine. *Mater. Today* **14**, 190–195.

Kearney J. B., Wheeler S. R., Estes P., Parente B., & Crews S. T. (2004). Gene expression profiling of the developing Drosophila CNS midline cells. *Dev. Biol*. **275**, 473-492.

Klämbt, C., Jacobs, J. R., & Goodman, C. S. (1991). The midline of the Drosophila central nervous system: A model for the genetic analysis of cell fate, cell migration, and growth cone guidance. *Cell* **64**, 801–815.

Kleve, C. D., Siler, D. A, Syed, S. K., & Eldon, E. D. (2006). Expression of 18-wheeler in the follicle cell epithelium affects cell migration and egg morphology in Drosophila. *Dev. Dynam.* **235**, 1953–1961.

Kobe, B. & Kajava, A. V. (2001). The leucine-rich repeat as a protein recognition motif. *Curr*. *Opin*. *Struc*. *Biol*. **11**, 725–732.

Kolesnikov, T. & Beckendorf, S. K. (2007). 18 wheeler regulates apical constriction of salivary gland cells via the Rho-GTPase-signaling pathway. *Dev. Biol.* **307**, 53–61.

Kurosaka, S. & Kashina, A. (2008). Cell biology of embryonic migration. *Birth Defects Res. C. Embryo Today* **84**, 102–122.

McIlroy, G., Foldi, I., Aurikko, J., Wentzell, J. S., Lim, M. A., Fenton, J. C., Gay, N. J., & Hidalgo, A. (2013). Toll-6 and Toll-7 function as neurotrophin receptors in the Drosophila central nervous system. *Nat*. *Neurosci*. **16**, 1248–1256.

Medzhitov, R. (2001). Toll-like receptors and innate immunity. *Nat*. *Rev*. *Immunol*. **1**, 135–145.

Medzhitov, R., Preston-Hurlburt, P., & Janeway, C. A. Jr. (1997). A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* **388**, 6–9.

Morisalo, D. & Anderson, K. V. (1995). Signaling Pathways that Establish the Dorsal-Ventral Pattern of the Drosophila Embryo. *Annu. Rev. Genet.* **29**, 371–399.

Muraki, K. & Tanigaki, K. (2015). Neuronal migration abnormalities and its possible implications for schizophrenia. *Front. Neurosci.* **9**, 1–10.

Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayasi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K., & Tabata, S. (1994). Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res.* **1,** 27–35.

Ooi, J. Y., Yagi, Y., Hu, X., & Ip, Y. T. (2002). The Drosophila Toll-9 activates a constitutive antimicrobial defense. *EMBO Rep.* **3**, 82–87.

Paré, A. C., Vichas, A., Fincher, C. T., Mirman, Z., Farrell, D. L., Mainieri, A., & Zallen, J. A. (2014). A positional Toll receptor code directs convergent extension in Drosophila. *Nature* **515**, 523–527.

Patel, N. H., Schafer, B., Goodman, C. S., & Holmgren, R. (1989). The role of segment polarity genes during Drosophila neurogenesis. *Gene. Dev.* **3**, 890–904.

Roark, M., Sturtevant, M. A., Emery, J., Vassin, H., Grell, E., & Bier, E. (1995). scratch, a pan-neural gene encoding a zinc finger protein related to snail, promotes neuronal development. *Gene. Dev.* **9**, 2384–2398.

Sarrazy, V., Vedrenne, N., Billet, F., Bordeau, N., Lepreux, S., Vital, A., Jauberteau, M. O., & Desmoulière, A. (2011). TLR4 signal transduction pathways neutralize the effect of Fas signals on glioblastoma cell proliferation and migration. *Cancer Lett.* **311**, 195–202.

Seppo, A., Matani, P., Sharrow, M., & Tiemeyer, M. (2003). Induction of neuron-specific glycosylation by Tollo/Toll-8, a Drosophila Toll-like receptor expressed in non-neural cells. *Development* **130**, 1439–1448.

Silverman, N., Paquette, N., & Aggarwal, K. (2009). Specificity and signaling in the Drosophila immune response. *Invertebr. Surv. J. (ISJ)* **6**, 163–174.

Stapleton, M., Liao, G., Brokstein, P., Hong, L., Carninci, P., Shiraki, T., Hayashizaki, Y., Champe, M., Pacleb, J., Wan, K., Yu, C., Carlson, J., George, R., Celniker, S., & Rubin, G. M. (2002). The Drosophila Gene Collection: Identification of putative full-length cDNAs for 70% of D. melanogaster genes. *Genome Res.* **12**, 1294–1300.

Takeda, K. & Akira, S. (2005). Toll-like receptors in innate immunity. *Int. Immunol.* **17**, 1–14.

Tessier-Lavigne, M., Placzek, M., Lumsden, A., Dodd, J., & Jessell, T. (1988). Chemotropic guidance of developing axons in the mammalian central nervous system. *Nature* **336**, 775–778.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E., & Rubin, G. M. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biol*. **3**, RESEARCH0088.

Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S. E., & Rubin, G. M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biol*. **8**, R145.

Trotta, T., Porro, C., Calvello, R., & Panaro, M. A. (2014). Biological role of Toll-like receptor-4 in the brain. *J. Neuroimmunol.* **268**, 1–12.

Truman, J. W., Schuppe, H., Shepherd, D., & Williams, D. W. (2004). Developmental architecture of adult-specific lineages in the ventral CNS of Drosophila. *Development* **131**, 5167–5184.

Valanne, S., Wang, J. H., & Rämet, M. (2011). The Drosophila Toll signaling pathway. *J. Immunol*. **186**, 649–656.

Van Aelst, L. & D'Souza-Schorey, C. (1997). Rho GTPases and signaling networks. *Gene. Dev*. **11**, 2295–2322.

Wang, J., Tao, Y., Reim, I., Gajewski, K., Frasch, M., & Schulz, R. A. (2005). Expression, regulation, and requirement of the Toll transmembrane protein during dorsal vessel formation in Drosophila melanogaster. *Mol. Cell. Biol*. **25**, 4200–4210.

Wheeler, S. R., Kearney, J. B., Guardiola, A. R., & Crews, S. T. (2006). Single-cell mapping of neural and glial gene expression in the developing Drosophila CNS midline cells. *Dev. Biol.* **294**, 509–524.

Wheeler S. R., Stagg S. B., Crews S. T. (2008). Multiple Notch signaling events control Drosophila CNS midline neurogenesis, gliogenesis and neuronal identity. *Development* **135**, 3071–3079.

Xiao H., Hrdlicka L. A., & Nambu J. R. (1996). Alternate functions of the single-minded and rhomboid genes in development of the Drosophila ventral neuroectoderm. *Mech. Dev*. **58**, 65-74.

Yamaguchi, H., Wyckoff, J., & Condeelis, J. (2005). Cell migration in tumors. *Curr. Opin. Cell Biol.* **17**, 559–564.

Yoon, H. J., Jeon, S. B., Koh, H. S., Song, J. Y., Kim, S. S., Kim, I. H., & Park, E. J. (2015). Distinctive responses of brain tumor cells to TLR2 ligands: *Glia* **63**, 894–905.

CHAPTER TWO: OPTIMIZATION OF A PROTOCOL TO USE FOR FACS AND SINGLE-CELL RNA SEQUENCING

INTRODUCTION

Neurogenesis is a critical aspect of development

Neurogenesis is the process by which neurons are generated from progenitors and multipotent, self-renewing neural stem cells (NSCs) throughout embryogenesis, and is critical for the proper development of the nervous system. Multiple signaling pathways, such as NOTCH, Wingless-related integration site (WNT), and Sonic Hedgehog (SHH) are required for newborn cells to properly differentiate and form the complex network of neuronal and glial populations in a mature brain (Jobe et al., 2012). During development, these cells must also migrate to their appropriate destinations, a process which is vital for the formation of the synaptic circuitry (Ghashghaei et al., 2007; Jobe et al., 2012). Defects in any of these processes can lead to severe developmental malformations such as schizencephaly (Spalice et al., 2009; Yakolev & Wadsworth, 1944), megalencephaly (Homem et al., 2015; Mirzaa & Poduri, 2014), and focal cortical dysplasias (Barkovitch et al., 2012; Kabat & Król, 2012). These structural deficits in the brain are characterized by, respectively, abnormal clefts, overgrowth of the brain, and disorganization of the cortex. There are many common symptoms associated with these conditions, including seizures, and poor cognitive and motor function. In severe cases, the prognosis is poor and survival into adulthood is rare. Understanding how neurons and glial properly differentiate and migrate during development will provide insight into how these disorders manifest, and hopefully ways to treat them.

Elucidating basic nervous system development at a single-cell level is crucial to answering key questions about how and why developmental processes are altered. By sequencing single cells, new roles for single genes or combinations of genes in neurodevelopmental disorders may be discovered, which would contribute to the understanding of these disorders as a whole. Despite the advances in knowledge regarding embryonic neurogenesis, some general questions remain. For instance, how is it that so few seemingly identical progenitors can give rise to such neuronal and glial diversity? What are the genetic switches that decide a cell's fate and ultimate destination in the brain? We may begin to find answers to these questions, as well as further define molecular mechanisms underlying embryonic neural development with the goal of developing therapies for certain neurodevelopmental disorders.

The *Drosophila* midline comprises a diverse group of cells during development

Drosophila embryogenesis is well characterized and occurs over an approximately 17 hour period after egg laying (AEL). During this time, the midline undergoes complex morphological changes, such as midline retraction, while cells are simultaneously differentiating into various subsets of neurons and glia. All midline cells are distinguished through expression of the transcription factor *single-minded* (*sim*), which is necessary for proper midline development and the expression of other midline-specific genes (Crews 1998; Fontana & Crews, 2012; Wheeler et al., 2006). About 3.5 - 4.5 hours AEL the midline is composed of sixteen seemingly similar cells per segment; by the end of development around 17 hours AEL, these cells are mature neurons and can be identified via differential gene expression (Wheeler et al., 2006) (**Table 2**). In the mature embryonic CNS, there are about 22 midline cells per segment,

which include: a dopaminergic H-cell interneuron and glutamatergic H-cell sib interneuron, two peptidergic MP1 neurons, three glutamatergic/octopaminergic mVUM motor neurons (innervate body wall muscles and the female reproductive system), and three GABAergic iVUM interneurons (Wheeler et al., 2006). This great diversity in neurons arises from a set of midline precursors (MPs), which each divide once during early development to generate two cells. MP1 divides to give rise to the MP1 neurons. MP3 generates H-cell and H-cell sib. MP4 divides into an iVUM and mVUM, as do MP5 and MP6. These divisions produce iVUM4-6 and mVUM4-6 (Wheeler et al., 2008). The MNB divides several times to generate approximately 8 neurons as progeny (Wheeler et al., 2006). As for midline glia (MG), there are two populations localized to either the anterior or posterior of embryonic segments (AMG and PMG, respectively). The PMG will undergo programmed cell death during late embryonic development; their roles in midline development are still unclear. The AMG will also undergo apoptosis, although there are survivors which ensheathe axon commissures (Fontana & Crews, 2012). Although all midline cells express *sim* as described, it is clear that considerable diversity exists in cell type, function, and fate.

Midline cell gene expression			
All midline cells	*sim*	MNB	*en*, *wor*
AMG	*runt, wrapper*	MP1; MP1 neurons	*lim3,odd*
H-cell	*gsb*, *ple*, *tup*	MP3	*gsb*
H-cell sib	*CG13565*, *gsb*, *VGlut*	MP4-6	*en*
iVUMs	*en*, *gad1*	PMG	*en*, *mas, wrapper*

Table 2: Midline cells can be identified by their expression of certain genes using either RNA probes or antibodies.

Previous methods to decipher the genetic differences between single cell types within the midline primarily relied on marking cells in fixed embryos using *in situ* hybridization and

antibody staining. While providing single-cell resolution, there are limits to the number of genetic markers which can be analyzed, making these procedures expensive and laborious. Further, these methods are limited in their ability to reveal novel genetic markers, which is largely accomplished through analysis of genetic mutants.

However, FACS and scRNA-seq vastly revolutionized this process. With FACS, an entire population of cells can be isolated in a high-throughput manner, and these thousands of cells can be prepared for further analysis. Utilizing scRNA-seq to interrogate the cell population would generate massive amounts of transcriptome data and provide a more complete picture of what is expressed in a given cell and when. The combined use of these techniques greatly improve the chance of uncovering novel genes involved in development, without sacrificing resolution. By exploiting the inherent genetic differences in these cells using flow cytometry and single-cell RNA sequencing, a greater understanding of the dynamic midline transcriptome can emerge and perhaps elucidate homologous processes in human neural development.

Fluorescence-activated cell sorting and RNA-sequencing are valuable tools for characterizing the transcriptome of heterogeneous cell populations

Fluorescence-activated cell sorting (FACS) is a specialized flow cytometry technique that can sort a heterogeneous mixture of cells into distinct populations on the basis of the cells' light scattering and fluorescent properties. Briefly, a single-cell suspension is prepared and loaded into a sorter. As cells emerge from the sorter, they pass single file across a small area illuminated by a laser beam, which detects the optical properties of the cells based on their measured fluorescence, scatter, and other preset criteria. A computer then converts the distinct optical properties of the cell into electrical pulses, thus providing a positive or negative charge to the isolated, single-cell droplets. The assigned charge deflects the droplet into the appropriate

collection tube (Herzenberg et al., 1976), thus isolating a cell population of interest from the original heterogeneous sample.

A wide variety of fluorophores can be utilized to tag cells of interest. In order to use fluorescent tags, the fluorophores must be biologically inert, fluorescently intense, and, if using more than one, exhibit little spectral overlap (Baumgarth & Roederer, 2000). FACS sorting has been successfully used with fluorophores in *Drosophila* to purify embryonic cell populations (Cumberledge & Krasnow, 1994; Fontana & Crews, 2012), hemocytes (Tirouvanziam et al., 2004), and follicle cells (Calvi & Lilly, 2004), among other populations. In these cases, it is typical to express GFP (or another fluorophore) in a tissue-specific manner using the UAS/Gal4 system. It is also possible to sort lineages of neuroblasts from tissue; in one experiment, *ase*-Gal4, a type I lineage specific marker, was used to drive the expression of nuclear GFP (UAS-stingerGFP) specifically in neuroblasts and to a lesser extent, their progeny. Cells were then sorted based on their size (differentiated cells were smaller than neuroblasts) and GFP intensity (differentiated cells gave off less signal than neuroblasts). Subsequent RNA-seq provided transcriptome data about this previously elusive set of cells, and led to the discovery that Klumpfuss can regulate self-renewal in neural stem cells (Berger et al., 2012).

Once a heterogeneous cell population is subdivided, it is possible to further characterize the sub-populations using techniques such as RNA-sequencing (RNA-seq). RNA-seq is a high throughput technology that takes advantage of next-generation sequencing to provide a snapshot of a cell population's transcriptional profile, or transcriptome. In this procedure, mRNA is isolated by its poly-A tail and primed for production of complementary DNA (cDNA) with a reverse transcriptase and primers, which can be sequence specific, or oligodT primers (which recognize the polyA tails of all mRNAs). The resulting cDNA libraries are then prepared for

high-throughput sequencing. The reads generated from sequencing are subsequently analyzed and mapped to a reference genome or transcriptome (Wang et al., 2009).

Taking the resolution of transcriptional analysis one step further is a recent advancement in sequencing technology called single-cell RNA sequencing (scRNA-seq), which can provide the expression profiles of individual cells within a population. This technology makes the characterization of a subpopulation of cells within a heterogeneous sample possible, thus providing a more detailed picture of the system as a whole (Wang et al., 2009). There are currently several different methods for isolating single cells, from micromanipulations to laser-capture microdissection (LCM) to microfluidic technology (Shapiro et al., 2013; Zong et al., 2012). As with any technique, there are benefits and limitations to each isolation method. For instance, a large benefit of LCM is that the spatial location of the cell sample is known, since tissue is selectively cut. However, it is low-throughput and may not be appropriate for subsequent transcriptome analysis, since it is difficult to capture all or most of a single cell without also collecting material from adjacent cells (Shapiro et al., 2013). Micromanipulations to isolate cells, such as through mouth pipetting or serial dilutions, are easy and inexpensive to perform. However, cells must be in suspension already, these techniques are low-throughput, and experimental error can be introduced through the misidentification of cells (Shapiro et al., 2013). Thus, the research question and subsequent experimental design must be considered carefully to ensure that the most suitable techniques are chosen prior to downstream sequencing.

To further elucidate the transcriptional profile of individual midline cells, FACS was used in conjunction with the Fluidigm C_1 Single-Cell Auto Prep System, which uses microfluidics technology to capture and process cells for sequencing. Essentially, a FACS-isolated suspension of midline cells is washed over the C_1 chip, which contains 96 wells to

capture single cells; these cells are then lysed, reverse transcribed, and amplified, all within the same capture chamber. A cDNA library can then be created and sequenced for each captured cell, revealing the expression profiles of individual cells, rather than a composite average (Egidio et al., 2014).

MATERIALS AND METHODS

Drosophila strains and genetics

The *elav-Gal4;UAS-2xEGFP* (*elav>2xEGFP*) strain generously donated by Daniel McKay at UNC-Chapel Hill was used for cell sorting experiments.

Collection and dissociation of embryos for cell sorting

elav>2xEGFP flies were bred in 6oz bottles containing cornmeal-based food at 25°C until approximately 50 bottles of flies were amassed. For embryo collections, about 1.5g of recently eclosed flies were transferred to cages and allowed to adapt for a few days. On collection days, 100mm grape juice agar plates streaked with yeast paste were cleared of older embryos once in the morning. Embryos were then collected for 2 hours on fresh yeast-streaked, grape juice agar plates and allowed to age an additional 14-16 hours after egg laying (AEL) at 25°C. After verifying stages collected, the embryos were dechorionated with 100% bleach for one minute, and extensively rinsed with water before being transferred to a 35mL glass Dounce homogenizer filled with 1X phosphate-buffered saline (PBS)+0.1% Triton-X-100 and allowed to settle to the bottom. Once embryos were settled, PBS+0.1% Triton-X-100 was carefully poured off and replaced with "hemolymph-like" (HL) buffer (25mM KCl, 90mM NaCl, 4.8mM $NaHCO_3$, 80mM d-glc, 5mM Trehalose, 5mM L-Gln, 10mM HEPES, final pH to 6.9) + 5mM EDTA (Salmand et al., 2011). For stage 16-17 embryos, mechanical disruption was achieved using nine strokes with a loose pestle and one final stroke with the tight pestle in the Dounce homogenizer. The resulting cell suspension was gravity filtered through Miracloth into 15mL culture tubes, filled to 10mL with HL + 5mM EDTA buffer and centrifuged at 100 x g for 10 minutes at 4°C. All but the last 1mL of supernatant was pipetted off and used to resuspend the cells by gentle

pipetting. The suspension was re-filtered though a 41μm mesh filter into clean 15mL culture tubes, filled to 10mL with HL + 5mM EDTA buffer, and centrifuged again at 100 x g for 10 minutes at 4°C. All but the last mL of supernatant was pipetted off and used to resuspend the cells, after which the suspension was filtered through a 20μm mesh filter. A 1mL aliquot of filtered suspension was reserved and stained with DAPI (1μL of 5mg/mL stock) to obtain a cell count with a hemacytometer. Based upon the cell counts, the remaining filtered cells were diluted in HL + 5mM EDTA buffer to $1x10^7$ cells/mL density and transferred to 5mL round-bottom polypropylene tube. At this point, cells were kept on ice in the dark for up to two hours until ready to sort. Before sorting, propidium iodide (PI) was added at a final concentration of 1μg/mL to serve as a live-dead cell marker.

Cell sorting

Cell sorting occurred on the same day approximately an hour after embryo dissociation on the Beckman Coulter MoFlo cell sorter, housed at the Flow Cytometry Core Facility of the UNC-Chapel Hill School of Medicine. Cell sorting was based on forward and side scatter properties, the presence of PI fluorescence which indicates dead and dying cells, and the intensity of GFP fluorescence. Using these properties for gating, cells were sorted into two receptacles containing either GFP^+ midline cells or GFP^- non-midline cells and placed on ice. After the sort, an aliquot of sorted cells was stained with DAPI and analyzed to verify the accuracy of the sorter.

Statistical analysis

The Student's t-test was used for statistical analysis in the graphs presented.

RESULTS

An optimized protocol for isolation of *Drosophila* midline cells for use with FACS and single-cell RNA sequencing

The goal of this experiment was to optimize a protocol for the generation of single-cell suspensions of midline cells from *Drosophila* embryos. These midline suspensions could then be used for scRNA-seq, in order to create a complete, single-cell resolution, transcriptional profile of *Drosophila* midline development. Here, *elav>2xEGFP* fly embryos were dissociated and filtered to obtain live, single cell suspensions, which were run through a FACS sorter to separate midline cells (GFP+) from non-midline cells (GFP-). From that point, the live, purified population of midline cells could be used for sequencing and analysis of the transcriptome using microfluidic technology.

The first step of this protocol is the isolation of live midline cells from embryos. In order to isolate midline cells using FACS, I used a fly line that employs the *UAS-Gal4* system to drive the expression of enhanced GFP (EGFP) in all *elav*-positive cells (*elav>2xEGFP*). In later stages of embryo development (stages 14-16), *elav* is expressed in the mature midline and brain.

Early versions of the original protocol for the dissociation of midline cells from embryos were not optimized for the formation of single-cell suspensions (Fontana & Crews, 2012). In these initial isolations, along with the presence of multi-cell clusters, many pieces of small and large cellular debris were present (defined as about <5μm and >5μm in diameter, respectively). Furthermore, many cells stained positive for propidium iodide (PI), indicating poor health or death. Finally, cell recovery was fairly low after performing all the filtering steps (conversation with Dr. Joseph Pearson, not published). All of these are concerns leading up to FACS sorting. If cells are clustered, even if they are GFP-positive, the sorter will discard the cells due to the larger size. Cellular debris can also interfere with sorting and prevent midline cells from being sorted

into the proper receptacle. Cells must also be alive and relatively healthy to undergo scRNA-seq. Although the chip used for subsequent sequencing only holds 96 cells, it is important to obtain several thousand, healthy cells to wash over the chip to increase the chances of capture. Thus, there was room for improvement in the protocol for cell dissociation.

The optimized protocol described produces a final single cell suspension with more cells, less debris, as well as fewer clusters and dead cells. Using this protocol, cells were run through a FACS sorter and several thousand healthy midline cells were successfully sorted from all other cells, demonstrating that this protocol can perform the same function as well as and better than the last.

Lowered trypsin and increased EDTA in dissociation buffer produces the fewest cell clusters

Leaving cells too long in any sort of protease is not conducive to cell health. In the original protocol, the dissociation buffer used was 0.5% trypsin, 0.2% EDTA in 1X PBS, which reduces cell clustering at the cost of cell viability. In order to increase viability and maintain reduced cell clustering, I compared cell viability and clustering between a control buffer (HL), a trypsin-only buffer (0.25% trypsin), an EDTA-only buffer (5mM), and a collagenase-only buffer (1mg/mL). Trypsin is a serine-threonine protease chosen for its ability to resuspend adherent cells and dissociate cells from tissue. Collagenase cleaves peptide bonds in collagen, and was selected to use since it digests connective tissue. EDTA, which chelates calcium and magnesium ions that can inhibit trypsin activity, also inhibits cadherin attachment between cells, and was thus selected to test for its efficacy in dissociation.

Aside from differences in dissociation buffers, cells were mechanically dissociated using the same methods described in the material and methods, and suspensions were analyzed for clusters using a hemocytometer. The buffer with the lowest number of cell clusters was the

EDTA-only buffer (5mM), followed by the trypsin buffer (0.25%) (**Figure 9A**). Compared to the 1X trypsin and control cell suspension, there were more clusters present in the 1X collagenase suspension (**Figure 9A**). The difference between the control and EDTA samples was not significant, though there was a significant difference in clustering between the EDTA and collagenase samples ($p<0.05$). In each of the samples, PI was used to identify dead cells; the fewest dead cells were seen in the control and EDTA samples, which were not statistically different. Thus, 1X EDTA in HL buffer (HL + 5mM EDTA) is the superior dissociation buffer tested, as it produces fewer cell clusters than the proteases and cell health is preserved.

Centrifuging at 100 x g for 10 minutes and pipetting off buffer is optimal for cell recovery

Another issue that was addressed from the original isolation protocol is cell recovery after centrifugation. Previously, cells were centrifuged after dissociation at 300 x g for ten minutes, after which the supernatant was decanted off. However, if cells are not all pelleted, there is a chance of losing many when pouring off liquid. To determine cell loss in the supernatant after centrifugation, samples were taken from the top (10mL mark), middle (5mL mark) and bottom of the culture tube after centrifugation and cells were counted on a hemocytometer (**Figure 9B**). Four samples per trial were used: one sample spun at 300 x g for 10 minutes, one sample spun at 300 x g for 20 minutes, one sample spun at 100 x g for 10 minutes, and one sample spun at 100 x g for 20 minutes. This trial was repeated four times. Centrifugation at a lower g was chosen as it should be gentler on the cells and cause less shearing, thus supporting cell health. However, a spin at 100 x g may not be sufficient to pellet all cells, so spin time was increased to 20 minutes, possibly resulting in fewer cells floating in the supernatant. It was found across trials that, although the majority of cells were localized to the bottom of the

tube after centrifugation, many cells were still floating in the top and middle of the sample regardless of spin time or velocity (**Figure 9C**). As a control to ensure that these cells are not prone to floating, a sample was spun down at 5000 x g for 10 minutes. As expected, virtually no cells were seen in the top or middle of the tube (data not shown). Comparing between the 10 and 20 minute spins, there was no significant difference in cell count between 300 x g at 10 and 20 minutes, or between 100 x g at 10 and 20 minutes. However, comparing spin velocities at 10 minutes, 100 x g was more efficient at pelleting cells to the bottom of the tube than 300 x g ($p<0.05$). At 10 minutes, the sample spun at 100 x g contained significantly fewer cells at the top of the tube than the sample spun at 300 x g ($p<0.01$) (**Figure 9C**). Based on this data, the protocol was modified to use a spin velocity of 100 x g for 10 minutes. In addition, instead of decanting the supernatant, it would be carefully pipetted off until only a milliliter of supernatant remained. This way, cells floating just above the pelleted cells could be recovered.

Filtering reduces the quantity of cellular debris present in samples

The next issue to address was that of cellular debris. Although it is difficult to remove cellular debris entirely, it is ideal to have a high cell to debris ratio before FACS sorting. To determine the cell to debris ratio present after centrifugation, the amount of large and small debris (>5μm, and <5μm respectively) was quantified after one spin at 100 x g for 10 minutes. It was found that cellular debris outnumbered cells at the top, middle, and bottom of the tube (**Figure 9D**). In order to decrease cellular debris, cells were filtered through Miracloth, a rayon-polyester filtration material, and 41μm and 21 μm mesh filters. Briefly, the procedure went as follows: dissociation, Miracloth filtration, centrifugation, 41μm filtration, centrifugation, and lastly 20μm filtration.

Gravity filtration through Miracloth immediately after dissociation successfully decreased the amount of large debris seen before centrifugation (data not shown). Next, I assessed the ability of the 41μm and 20μm mesh filters to decrease cellular debris between rounds of washes and centrifugation. For this, resuspended cells were gravity filtered through a 41μm mesh filter after the first wash and centrifugation, and then filtered again through a 20μm mesh filter after a second wash and centrifugation. Immediately after the first and second spins, a small aliquot of sample was taken to assess cell to debris ratio. After the first spin, a large amount of debris was present along with the cell pellet. Some cells and debris still lingered at the top and middle of the tube as well (**Figure 9E**). After a second spin, which was filtered through the 41μm mesh filter prior to centrifugation, considerably less debris was present at all points sampled. In addition, cell recovery after washing a second time was excellent and the ratio of cells:debris improved per 20nL sampled (**Figure 9E**). The data indicate that filtering before washing is beneficial for decreasing the amount of debris present, and that pipetting off the supernatant rather than decanting is aiding in cell recovery. The protocol was then modified to include all of the filtering steps detailed. To summarize the protocol:

1. Dissociate cells from tissue in HL + 5mM EDTA
2. Gravity filter through Miracloth
3. Wash/centrifuge at 100 x g for 10 minutes
4. Pipette off all but last mL of supernatant, gently pipette to resuspend pellet
5. Gravity filter through 41μm mesh filter into clean tube
6. Wash/centrifuge at 100 x g for 10 minutes
7. Pipette off all but last mL of supernatant, gently pipette to resuspend pellet
8. Gravity filter though 20μm mesh filter into clean tube

Two washes is sufficient for reducing debris and recovering cells

Since performing two washes was helpful in reducing the amount of debris present, it was thought that perhaps additional washes could be even more beneficial. Up to four washes, at 100 x g for 10 minutes, were conducted on samples and aliquots were taken from the bottom of the tube for counting. After the first spin, the same pattern is seen where the amount of debris present outnumbers the quantity of cells per 4nL (about 1:2.3 cells:debris) (**Figure 9F**). After the second wash, the ratio of cells:debris improves to about 1:1.4, supporting the results found previously (**Figure 9F**). However, after the third and fourth spins, cell recovery decreased without a proportional reduction in debris (about 1:4 and 1:2, respectively) (**Figure 9F**). The minimal decline in debris observed after more spins did not merit the continual loss of midline cells. Since more than two washes and centrifugation cycles is detrimental to total cell recovery, the protocol remained unchanged in that two washes and centrifugation cycles were used throughout the course of the protocol.

The optimized protocol had a final sample with adequate healthy cell recovery, fewer debris pieces, and little cell clustering

In order to determine the final proficiency of the protocol with the given changes, this procedure was run in its entirety using the findings from the optimization experiments, and the general time course of cell loss was plotted. Applying the changes to the protocol (summarized in **Table 3**), from beginning to end resulted in a cell recovery of approximately 30% (**Figure 9G**). Further, few cells stained positive for PI, virtually no cells clustered, and the amount of debris in the final sample decreased (data not shown). In conclusion, the updated protocol produces a cell suspension suitable for running through a FACS sorter, thus ensuring that healthy midline cells can be isolated.

Summary of protocol optimization	
Dounce strokes	9X with loose pestle, then 1X with tight pestle
Anti-clustering agent	5mM EDTA in HL buffer
Filtration	Miracloth, 41μm and 20 μm mesh filters
Spin velocity	100 x g
Spin times	10 minutes per wash
Washes	2
Supernatant	Pipette off all but the last mL of supernatant after a wash

Table 3: Summary of optimized protocol.

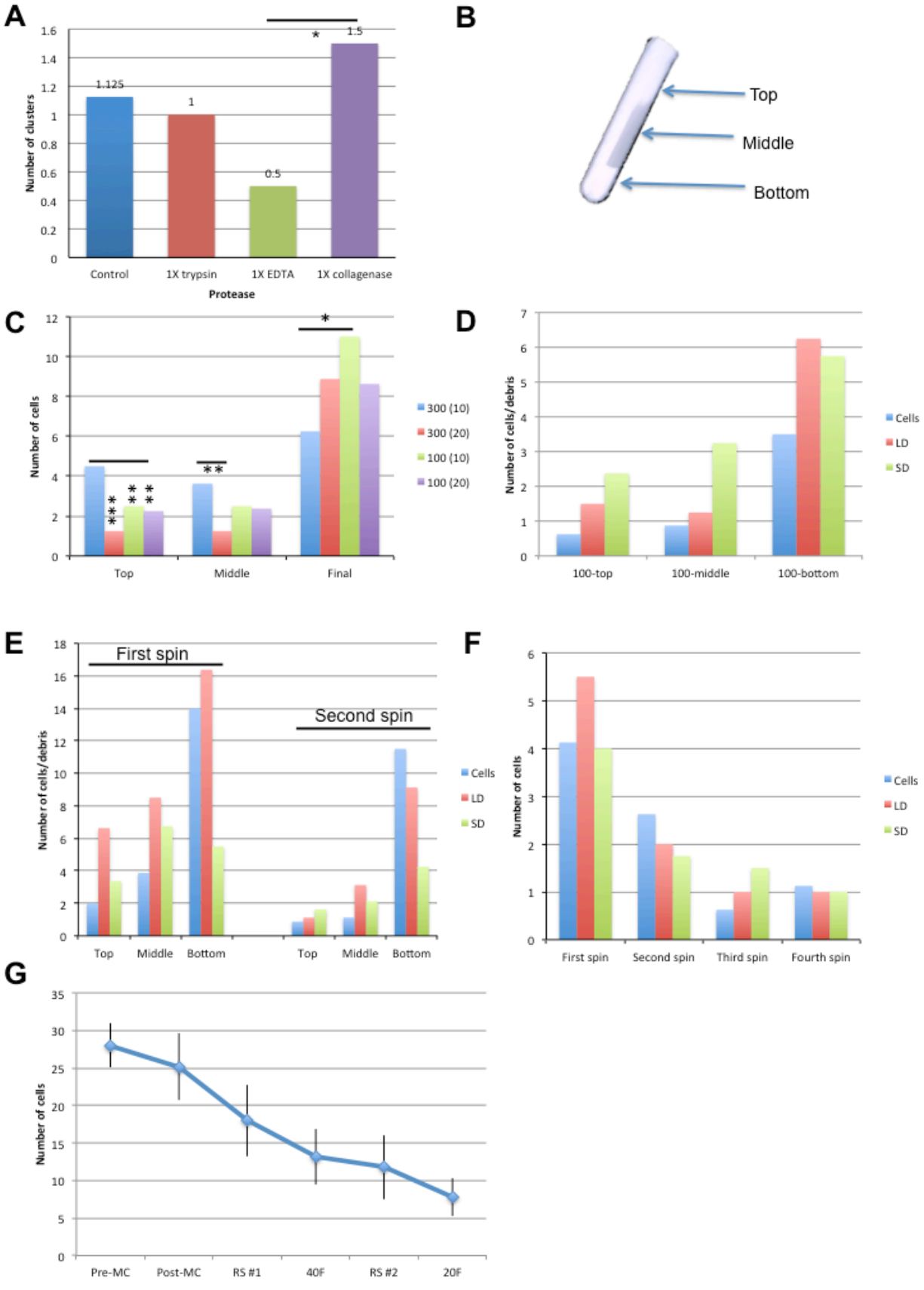
A
Number of clusters
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1.125
1
0.5
1.5
*
Control
1X trypsin
1X EDTA
1X collagenase
Protease
B
Top
Middle
Bottom
C
Number of cells
12
10
8
6
4
2
0
*
**
***
***
Top
Middle
Final
300 (10)
300 (20)
100 (10)
100 (20)
D
Number of cells/debris
7
6
5
4
3
2
1
0
100-top
100-middle
100-bottom
Cells
LD
SD
E
Number of cells/debris
18
16
14
12
10
8
6
4
2
0
First spin
Second spin
Top
Middle
Bottom
Top
Middle
Bottom
Cells
LD
SD
F
Number of cells
6
5
4
3
2
1
0
First spin
Second spin
Third spin
Fourth spin
Cells
LD
SD
G
Number of cells
35
30
25
20
15
10
5
0
Pre-MC
Post-MC
RS #1
40F
RS #2
20F

Figure 9: Quantification of different optimization parameters for tissue dissociation and obtaining a single cell suspension for FACS sorting. (A) The number of cell clusters per 4nL (y-axis) was counted for samples using a different anti-clustering agent (x-axis). The use of 1X EDTA in the dissociation buffer produced the fewest clusters observed between control, 1X trypsin, and 1X collagenase. Cells were counted on a hemocytomer and the number was averaged among 4nL squares. N=16 squares per sample, $p<0.05$*. (B) Representation of top, middle, and bottom of a tube from which cell samples were taken for counting. The top indicates a sample taken from the 10mL mark, middle indicates a sample taken from the 5mL mark, and bottom refers to the last mL. (C) Number of cells counted (y-axis) at different spin velocities and times from the top, middle, and bottom of a sample tube (x-axis). Blue bars show data for the 300 x g for 10 minutes spin, red bars for 300 x g for 20 minutes spin, green bars for the 100 x g for 10 minutes spin, and purple bars for the 100 x g for 20 minutes spin. Cells were still present in the top and middle of all samples after spinning, but the most cells per 4nL square were counted in the 100 x g sample spun for 10 minutes. N = 16 squares per sample, $p<0.05$*, $p<0.01$**, $p<0.001$***. (D) The quantity of cellular debris and cells (blue bar) was counted per 4nL at the top, middle, and bottom of the tube. Large debris (LD, red bar) was defined as debris larger than 5μm in diameter, while small debris (SD, green bar) was defined as debris smaller than 5μm in diameter. After a single spin at 100 x g for 10 minutes, most debris was located at the bottom of the tube and the cell:debris ratio was not ideal. N= 16 squares per sample. (E) After a second spin at 100 x g for 10 minutes, the quantity of cellular debris and cells (blue bar) was counted per 4nL at the top, middle, and bottom of the tube. LD (red bar) and SD (green bar) were reduced overall after the second wash and the cell:debris ratio improved. N = 16 squares per sample. (F) Additional washes were conducted to determine if debris could be decreased further without sacrificing cell recovery. Cells (blue bar), LD (red bar), and SD (green bar) were counted from the bottom of the tube after each wash. The third and fourth wash decreased the amount of cellular debris, but at the expense of cell recovery. Two washes were deemed optimal for both reducing debris and recovering an adequate amount of cells. N = 16 squares per sample. (G) General time course of cell loss throughout the tissue dissociation procedure following the optimized protocol. Pre-MC = pre-Miracloth filtering, post-MC = post-Miracloth filtering before washing, RS #1 = cell resuspension #1 after the first wash, 40F = post 41μm mesh filtering, RS #2 = cell resuspension #2 after the second wash, and 20F = post 20μm mesh filtering, which is the final suspension that would be taken to the FACS sorter. Approximately 30% of healthy cells were recovered from pre-MC to 20F.

Midline cells were isolated from a single-cell suspension using FACS

Using the optimized protocol, 16-18 hour old *elav>2xEGFP* embryos were dissociated and processed to obtain a single-cell suspension. The suspension was diluted to a density of $1x10^7$ cells/mL and kept on ice in the dark until ready to sort later that same day. PI was added immediately before sorting to serve as a live-dead marker. After gating for fluorescence from the control samples (w^{1118} with and without PI added, *elav>2xEGFP* without PI), the experimental sample (*elav>2xEGFP* with PI) was run through the Beckman Coulter MoFlo. For the first run, the machine was set to sort through 50,000 cells to isolate GFP-positive, PI-negative midline cells into a receptacle tube (**Figure 10A-B**). Approximately 2500 cells of interest were isolated.

Immediately following, the midline cells were sorted once again to verify enrichment and cell count. When set to the same parameters as the previously sorted cells, the MoFlo only recovered about 1600 cells, 700 of which were GFP-positive, PI-negative. The population was enriched for GFP-positive cells, as expected (**Figure 10C**). When an aliquot from the first GFP-positive sort and second GFP-positive re-sort was stained with DAPI and manually counted, it was found that the cell count was consistently about 10-15% of the count suggested by the sorter. All cells collected appeared relatively healthy. This is useful information to know before performing RNA-seq so enough cells can be isolated to ensure a successful experiment.

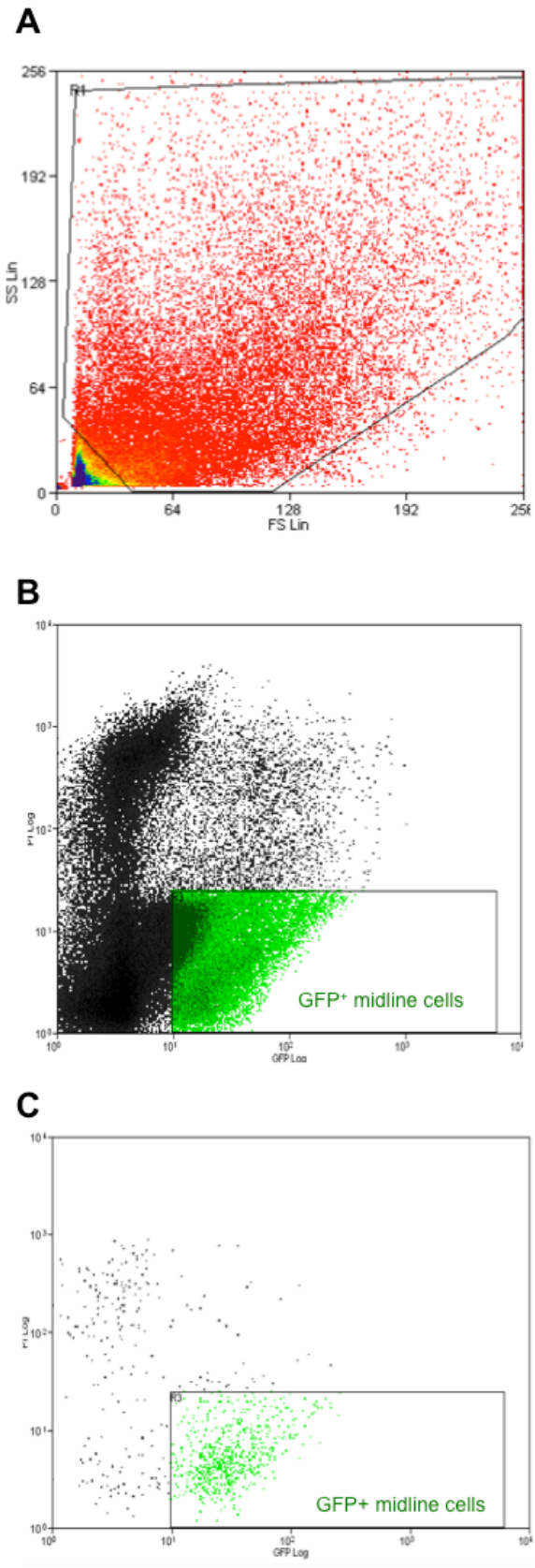

Figure 10: Midline cells were sorted from a heterogeneous population of cells using FACS. (A) Side scatter (SS, proportional to cell granularity) versus forward scatter (FS, correlates with cell size) was plotted for 50,000 cells. The cells within the polygon drawn are the cells that underwent further analysis. (B) Log of PI fluorescence versus log of GFP fluorescence for the cells within the gated population. The rectangle encloses GFP-positive, PI-negative midline cells. This is the population of cells that were isolated for a subsequent sort. (C) Log of PI fluorescence versus log of GFP fluorescence for the population of midline cells from (B) after a second sort. Although some cells died between runs (high PI fluorescence), many midline cells could be recovered (rectangle). This shows that the protocol optimized can be used to successfully perform a FACS experiment to isolate cell populations of interest.

DISCUSSION

Single cell sequencing is a powerful tool with the potential to provide precise resolution into the dynamics of cellular processes at any given time. With this technology, we can pinpoint cellular differences and obtain a better understanding of how the cell functions within a given environment, such as the midline. Identifying these cellular differences and functions can lead to the elucidation of certain developmental processes, such as neuronal differentiation, and perhaps identify novel targets for the treatment of disorders. In order to take advantage of this technology, it is necessary to use a protocol developed specifically for isolation of midline cells from *Drosophila* embryos. In this chapter, I describe the optimization of a protocol used to obtain the highest possible number of healthy cells in a single cell suspension with the lowest quantity of cellular debris, a situation ideal for subsequent FACS sorting. This protocol successfully isolates midline cells from non-midline cells, and acts as a proof of principle that live, healthy cells can be recovered for later RNA-seq and analysis.

The original protocol required optimizations to decrease cell clustering and cellular debris, while maintaining cell viability. Originally, trypsin-EDTA was used to break clusters of cells apart, and then neutralized with serum. However, incubating cells with trypsin for too long can damage cell membranes, endangering cell health, or possibly kill the cells, which is detrimental for FACS sorting. EDTA, a divalent cation chelator, is generally added to trypsin stocks as it reduces the presence of free Mg^{2+} and Ca^{2+}, both of which inhibit trypsin activity. However, this reduced availability of calcium inhibits cadherin-mediated cell-cell adhesion. In summary, EDTA promotes activity of trypsin and decreases cell-cell adhesion, and thus provides a more gentle method for reducing cell clustering than actively cleaving proteins. This feature made it an attractive anti-clustering agent to use in the dissociation buffer. Indeed, most cells

appeared healthy after prolonged contact with EDTA. This eliminated the need for trypsin and the serum neutralization step, making the protocol more streamlined. Because live cells are needed for FACS sorting, this is ideal since as little time as possible should pass between tissue dissociation and sorting.

Centrifugation allows for increased cell recovery, but high speeds risk isolation of cellular debris. The original protocol also called for centrifuging samples twice at 300 x g for 10 minutes in order to wash cells. During optimization, I found that spinning cells for 20 minutes rather than 10 minutes made no significant difference in increased cell recovery, indicating that cells are moving towards the bottom of the tube within the first 10 minutes of spinning. Once again, the decreased wash time means the protocol can be completed quicker and cells can be sorted sooner. Further, spinning at 100 x g was as efficient at spinning at 300 x g in reference to cell recovery. In addition, fewer dead cells and generally less debris were seen in the 100 x g samples than in the 300 x g samples (data not shown). The increased debris and dead cell count observed in the 300 x g samples is likely due to greater cell shearing. In conclusion, the optimal wash time is 10 minutes while centrifuging at 100 x g.

Cell filtration is used to remove larger pieces of cellular debris. The original protocol utilized a 41μm mesh filter after the first wash to manually push cells through the filter, thus isolating whole cells from larger pieces of debris. Forcing cells through a filter is stressful for cells, and may also contribute to the quantity of debris if cells break apart under the pressure or encounter a clog. To minimize the presence of large debris, which would cause clogs, Miracloth was used to gravity filter out large pieces of debris from the samples prior to the first wash. Thus, when the 41μm mesh filter was used after the first wash, cells flowed easily through with no need for applied manual pressure. This is presumably because the larger pieces of debris, which

may have normally clogged the 41μm filter, were removed before the wash. Following the second wash, cells were filtered through a mesh filter, and again, flowed easily through. Although cell recovery decreased after each filtering step, the amount of debris present was reduced and fewer dead cells were present at the end of the protocol. Overall, the preservation of cell health and reduction of cellular debris outweighed the small reduction in cell numbers, and so the Miracloth and 20μm filter were added to the protocol.

Washes served to remove cellular debris, but too many washes decrease live cell yields. The original protocol called for two washes and it was determined that two washes was sufficient for removing debris, as well as recovering as many cells as possible. The debris removed by performing an extra third or fourth wash was negligible, and there was a positive correlation between number of washes and cells lost. Thus, it was decided that the loss in cells was not worth the slight decrease seen in the amount of cellular debris, and so only two washes were utilized in the final protocol.

While many cells are pelleted after centrifugation, a percentage remains floating in the supernatant. As shown in Figure 9, there are actually cells present throughout the tube after spinning, especially towards the bottom just above the cell pellet. Originally, after each wash, the supernatant was decanted off, resulting in considerable cell loss. Carefully pipetting off all but the last milliliter and resuspending from that point saves these cells from being discarded and contributes to overall cell recovery. It is important to recover as many cells as possible, especially because the midline cell population is small relative to non-midline cells. The more cells obtained, the more midline cells can be isolated via FACS. Though more time-consuming than decanting, considerably more cells can be recovered for downstream analysis by pipetting off the supernatant, and so this step was added to the finalized protocol.

The FACS data reveals that this protocol can yield a relatively healthy, pure population of midline cells from a heterogeneous cell suspension. However, one drawback is that less midline cells were recovered from a pure sample of previously sorted cells. However, this re-sort would not generally be performed during non-optimization procedures. In any case, there are several possibilities to explain the observed loss in midline cells. Optimally, when cells are sorted and immediately re-sorted, the expectation is that the majority of cells will be recovered with little cell death in between sorts. In this case, it is possible that between sorts, some GFP-positive midline cells died, and thus pushed them past the gate for PI fluorescence, leading the sorter to discard them. It is also possible that some cells clustered together between runs, which would also lead the sorter to discard them. In addition, some cells may have stuck to the collection tube after the first sort and thus were never run through the second sort. This is not an exhaustive list of all possible reasons why the cell count was lower on the second run, and it may be difficult to determine the precise cause (or causes) of this occurrence. For instance, one could examine the original midline sort tube under a microscope to determine if cells are stuck to the side. One could also perform a time series of cell death; take small aliquots of sorted cells after a predetermined amount of time (every ten minutes, for example), and count the number of cells with PI fluorescence. This would provide some insight as to why the re-sort results in fewer midline cells. Although cells were lost between sorts (again, re-sorting is used to validate that a pure population can be collected and will likely not be performed before RNA-seq), it does reveal important information. Immediately after sorting, it is critical to count cells manually, assess cell health, and underestimate the quantity collected to ensure that a large enough population is collected for sequencing.

Although not performed here, the next step is to prepare the isolated midline cells for single-cell RNA-seq using the Fluidigm C_1 system. Using the previous RNA-seq technology without single cell analysis, the transcriptome of the midline could be analyzed at any given developmental time period, and the identity and expression levels of activated genes could be determined. However, it is difficult to determine from this dataset exactly which cells in a population are expressing a given gene, because expression levels are averaged across the entire heterogeneous sample. For instance, *wrapper* is expressed in midline glia, but not neurons; RNA-seq would simply indicate that *wrapper* is expressed in the midline. Further analyses would be necessary to isolate specifically which cells are *wrapper*-positive. Thus, these RNA-seq experiments only provide partial information about the molecular state of the system studied (Shapiro et al., 2013). Using scRNA-seq technology, the transcriptomes of individual midline cells can be analyzed and can reveal insight into differences between cell types as well as into the regulation of differentiating neurons and glia. This is by no means the limit. Using this optimized protocol, different midline cell types can be fluorescently tagged and isolated via FACS. This would allow for the purification and sequencing of subsets of midline neurons and glia. The knowledge gained from these experiments would contribute greatly to the understanding of the regulation of nervous system development, the factors underlying cell differentiation, and the differences between cell types.

Broadening this experiment, once the wild-type midline transcriptome has been thoroughly characterized, it would be interesting to sort cells from mutant flies to see if genes of interest (such as *18w*, previous chapter) are affecting the regulation of midline development and how. For instance, if *18w* is deleted, other genes may be up- or downregulated. This experimental design would allow for the rapid identification of multiple genes which could then

be studied further to determine a mechanism or pathway for *18w* action. These experiments would contribute to the knowledge about nervous system development and might suggest new genes or pathways to target to develop therapies for neurodevelopmental disorders.

REFERENCES

Barkovich, A. J., Guerrini, R., Kuzniecky, R. I., Jackson, G. D., & Dobyns, W. B. (2012). A developmental and genetic classification for malformations of cortical development: update 2012. *Brain* **135**, 1348–1369.

Baumgarth, N. & Roederer, M. (2000). A practical approach to multicolor flow cytometry for immunophenotyping. *J. Immunol. Methods* **243**, 77-97.

Berger, C., Harzer, H., Burkard, T. R., Steinmann, J., van der Horst, S., Laurenson, A. S., Novatchkova, M., Reichert, H., & Knoblich, J. A. (2012). FACS Purification and Transcriptome Analysis of *Drosophila* Neural Stem Cells Reveals a Role for Klumpfuss in Self-Renewal. *Cell Rep*. **2**, 407–418.

Calvi, B.R. & Lilly, M.A. (2004). Fluorescent BrdU labeling and nuclear flow sorting of the *Drosophila* ovary. *Methods Mol. Biol*. **247**, 203–213.

Crews, S. T. (1998). Control of cell lineage-specific development and transcription by bHLH-PAS proteins. *Genes Dev*. **12**, 607–620.

Cumberledge, S. and Krasnow, M.A. (1994). Preparation and analysis of pure cell populations from *Drosophila*. *Methods Cell Biol*. **44**, 143–159.

Egidio, C., Ooi, A., Holcomb, I., Ruff, D., Wang, J., & Ramakrishnan, R. (2014). A method for detecting protein expression in single cells using the C_1^{TM} Single-Cell Auto Prep System. Poster presented at AAI Annual Meeting in Pittsburgh, PA.

Fontana, J. R. & Crews, S. T. (2012). Transcriptome analysis of Drosophila CNS midline cells reveals diverse peptidergic properties and a role for castor in neuronal differentiation. *Dev. Biol.* **372**, 131–142.

Ghashghaei, H. T., Lai, C., & Anton, E. S. (2007). Neuronal migration in the adult brain: are we there yet? *Nat. Rev. Neurosci*. **8**, 141–151.

Herzenberg, L. A., Sweet, R. G., & Herzenberg, L. A. (1976). Fluorescence-activated cell sorting. *Sci. Am.* **234**, 108–117.

Homem, C. C. F., Repic, M., & Knoblich, J. A. (2015). Proliferation control in neural stem and progenitor cells. *Nat. Rev. Neurosci.* **16**, 647–659.

Jobe, E. M., McQuate, A. L., & Zhao, X. (2012). Crosstalk among epigenetic pathways regulates neurogenesis. *Front. Neurosci*. **6**, 59.

Kabat, J. & Król, P. (2012). Focal cortical dysplasia – review. *Pol. J. Radiol.* **77**, 35–43.

Mirzaa, G. M. & Poduri, A. (2014). Megalencephaly and hemimegalencephaly: Breakthroughs in molecular etiology. *Am. J. Med. Genet. Part C.* **166C**, 156–172.

Salmand, P. A., Iché-Torres, M., & Perrin, L. (2011). Tissue-specific cell sorting from Drosophila embryos: Application to gene expression analysis. *Fly* **5**, 261–265.

Shapiro, E., Biezuner, T., & Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630.

Spalice, A., Parisi, P., Nicita, F., Pizzardi, G., Del Balzo, F. and Iannetti, P. (2009), Neuronal migration disorders: clinical, neuroradiologic and genetics aspects. *Acta Paediatr*. **98**, 421–433.

Tirouvanziam, R., Davidson, C.J., Lipsick, J.S., and Herzenberg, L.A. (2004). Fluorescence-activated cell sorting (FACS) of *Drosophila* hemocytes reveals important functional similarities to mammalian leukocytes. *Proc. Natl. Acad. Sci. USA* **101**, 2912–2917.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Gen.* **10**, 57–63.

Wheeler, S. R., Kearney, J. B., Guardiola, A. R., & Crews, S. T. (2006). Single-cell mapping of neural and glial gene expression in the developing Drosophila CNS midline cells. *Dev. Biol.* **294**, 509–524.

Wheeler S. R., Stagg S. B., Crews S. T. (2008). Multiple Notch signaling events control Drosophila CNS midline neurogenesis, gliogenesis and neuronal identity. *Development* **135**, 3071–3079.

Yakolev, P. I. & Wadsworth, R. C. (1944). Schizencephalies: a study of congenital clefts in the cerebral mantle. I. Clefts with fused lips. *J. Neuropathol. Exp. Neurol.* **5**, 116–130.

Zong, C., Lu, S., Chapman, A. R., & Xie, X. S. (2012). Genome-wide detection of single nucleotide and copy number variations of a single human cell. *Science* **338**, 1622–1626.