Neuron Identity Problem: From Genes to Function

Citation

Pandey, Shristi N/A. 2020. Neuron Identity Problem: From Genes to Function. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Neuron Identity Problem: From Genes to Function

A dissertation presented

by

Shristi Pandey

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of Biology

Harvard University

Cambridge, Massachusetts

October 2019

Dissertation Advisor: Professor Alexander F. Schier Shristi Pandey

Neuron Identity Problem: From Genes to Function

Abstract

Cellular specialization is one of the key features of the vertebrate nervous system, which is composed of thousands to billions of neurons that that have distinct molecular, morphological, electrophysiological and connectivity properties. Deconstructing the nervous system into its constituent parts is one of the fundamental steps to understanding how it generates behavior and how its dysfunction might lead to neuropsychiatric disorders. To tackle this problem, we generated comprehensive cell type profiles of two regions of the zebrafish brain: 1) the habenula and 2) the telencephalon and sought to connect these molecular profiles to functional properties of said brain regions.

To generate comprehensive cell type maps of these regions, we produced and analyzed single-cell RNAseq data from the habenula and telencephalon. Using a two-time point single cell analysis of each region combined with machine learning approaches, we investigated questions on the maintenance and divergence of cell type identities between a developing animal and a mature adult. We found that the habenular cell types were largely maintained between larva and adult. On the other hand, cell types in the telencephalon displayed significant divergences between developing larva and mature adult.

To interrogate the functional roles of cell types, we combined molecular profiling approaches with *in situ* hybridization and computational image morphing to generate detailed spatial atlases for both regions of the brain. Using the spatial atlas of the habenula, we identified a population in the ventrolateral habenula that is activated by inescapable aversive environmental stimuli.

To further understand the functional role of the cell types and marker genes, we generated mutants of cell type specific marker genes and assayed them for brain activity and behavior. We discovered that all the habenular mutants displayed aberrant locomotor activity and one displayed a reduction in arousal threshold, establishing the habenula as an important regulator of activity and perhaps even sleep.

Finally, we utilized the cell type landscape of the forebrain to understand the molecular basis of a brain activity and behavioral phenotype in animals with loss of a schizophrenia associated gene *znf536*. Together, these observations demonstrate how the deconstruction of the brain into its constituent cell types can spur further discoveries of the role of cell types in generating animal behavior in health and disease.

Table of Contents

Acknowledgements

Several people contributed in many meaningful ways to the successful completion of my PhD. To them, I owe my deepest gratitude.

First, to my undergraduate advisor Philip Hanawalt who took a chance on me when I was a naïve freshman in college and instilled in me the confidence to move forward in science. My scientific journey began in his lab where I ran countless radiolabeled gels and was regaled by many fascinating stories from his own scientific journey. The third floor of Herrin Hall where Phil started his lab in 1961 will remain close to my heart.

To my PhD advisor, Alexander F. Schier for his wisdom, guidance, support, patience, and advice throughout my PhD. In addition to assembling an incredibly resourceful and diverse lab where I could grow as a scientist, Alex gave me an enormous amount of independence as a graduate student and provided many avenues for my growth. Through Alex's mentorship, I learned more than I had hoped for going into graduate school.

To my thesis advisory committee, Joshua Sanes, Jeffrey Macklis, Nao Uchida for their ideas and input over the years in improving my work and for their support and encouragement throughout my graduate career. Thank you to Catherine Dulac for joining in at the end.

To all members of the Schier lab, past and present for making the Schier lab a special place for science: Annika Nichols, Constance Richter, Vassilis Bitsikas, Amelia Hidalgo, William Joo, Joo Won Choi, Mehdi Goudarzi, Nate Lord, Adam Carte, Jeffrey Farrell, Yiqun Wang, Dave Dingal, Philip Abitua, Jessica Leslie, Bushra Raj, Tessa Montague, Martin Haesemeyer, Andrea Pauli, James Gagnon, Michal Rabani, Owen Randlett, Maxwell Shafer, Summer Thyme, Jenna Sternberg, Kathryn Berg, Megan Norris, Alyson Ramirez, Laila Akhmetova, Guo-Liang Chew, Alix Lacoste, Katherine Rogers, Jennifer Li, Drew Robson, Lindsey Pieper, Madalena Pinto, and Steve Zimmerman. Special thanks to Jeffrey Farrell and Martin Haesemayer who were always there to answer my questions when I needed.

To Constance Richter for bringing me optimism and encouragement when I needed it.

To Summer Thyme with whom I shared animated scientific and non-scientific discussions over many Indian buffets and late nights in lab. She has been a great friend, mentor and cheerleader over the years.

To Karthik Shekhar and Kristian Herrera for being great collaborators. I learned a lot from both of them.

To Jessica Miller, Karen Hurley and Brittany Hughes for taking such good care of all of the fish so that I could focus on the science.

To the Harvard Evening Van Crew for getting this night-owl home at odd hours of the night.

To my family and friends both in US and in Nepal for their love, support and curiosity in what I do. Special shout-out to Aditi Adhikari, for always being there and always checking in.

To the Sisteines: Stephanie, Haneui, Jenelle, Denise, Brenda and Vika, an amazing group of friends, who were there with me through all the ups, downs and in-betweens of graduate school life.

To Mr. and Mrs. McCaw, without whose generosity, a kid from Kathmandu would have never dreamt of going to a place like Stanford.

To my brother, whose jests and quips make me happy.

To my parents who saw me leave home for the first time about 10 years ago but remain close to my heart in all that I do. They made enormous sacrifices in their lives to give me the best possible future they could have hoped for. For this, I will remain eternally grateful.

To Sunil, my best friend and my biggest cheerleader, through ups and downs. I am lucky to have you in my life.

To my grandmother, who I miss dearly. I look forward to seeing you again.

Finally to NPR - for keeping me company through the humdrum of (certain types of) benchwork.

List of Figures

Appendix

List of Tables

Abbreviations

scRNAseq	Single Cell RNAseq
dHb	Dorsal Habenula
vHb	Ventral Habenula
Aoc1/Dao	Amine oxidase Copper Containing 1/ Diamine oxidase
Gpr139	G Protein-Coupled Receptor 139
Nts	Neurotensin
Ntsr1	Neurotensin Receptor 1
RNA-FISH	RNA Flourescent In Situ Hybridization
GWAS	Genome Wide Association Studies
Seq-FISH	Sequential Flourescent In Situ Hybridization
MERFISH	Multiplexed error-robust fluorescence in situ hybridization
IPN	Interpeduncular Nucleus

Chapter 1: Introduction

1.1 Classification of Neuronal Cell Types

Cell types are fundamental units of biological systems. In fact, the major driver for the evolution of multicellular life is thought to be the advantage that is conferred by specializing cell types to execute specific tasks (Arendt, 2008). Such a division of labor is a particularly prominent feature of the nervous system which has long been appreciated for the immense diversity of its cell types (Bota and Swanson, 2007; Luo et al., 2008). Neuroscientists have long been fascinated by this diversity of cell types in the brain. The concept that specific cell types with unique characteristics form building blocks of the nervous system dates back to the time of Santiago Ramon y Cajal and his contemporaries during the late 1800s (Ramon y Cajal, 1899). Using Camillo Golgi's 'black reaction' to stain sparse cell populations in histological sections, Cajal and his contemporaries identified a diversity of cell types based on anatomical features. In addition, they described the variation in these cell types across different brain regions and the conservation of said types across species.

Since then, other modalities have been added to the repertoire of cellular descriptors such as gene expression, projection patterns and functional/electrophysiological properties. However, until recently, these descriptors were still explored using low-throughput histological methods (Bota and Swanson, 2007; Masland, 2004) and remained severely underpowered. These approaches are not scalable to systematically and reproducibly classify majority of the cell types in the brain. The cellular complexity of the nervous system relative to these traditional techniques has impeded the development of a cell inventory of the brain for a long time.

High-throughput methods such as single cell RNA-seq have provided avenues to collect comprehensive information about individual cells in the brain to stratify neurons into cell types based on a large number of features or genes. In contrast to imaging approaches, this approach allows the profiling of the complete gene expression state of a particular neuron, which in theory should encode morphological, connectivity and functional features of the cell given this profiling is done in a dynamic manner so as to capture the development history of a particular neuron.

While these approaches improve the scalability of the age-old problem of cell type classification in the brain, the question of what constitutes a cell type is still not without contention. In this thesis, I explore some of the various points surrounding the problem of cell type classification in the brain. In this thesis, a cell type is defined as a shared, stable molecular ‘ground state’ that broadly dictates the functional capacities of a particular neuron (Fishell and Heintz, 2013). While individual members of a cell type may differ in the exact profile of genes that they express depending on context and activity, all cells belonging to that type should express a stable common suite of genes. In the sections below, I will describe the various facets of this problem in greater detail and the features that should be considered when tackling the problem of defining cell type identities in the brain.

1.2 The Multi-layered Problem of Defining Cell Types in the Brain

While single-cell RNAseq provides a unique framework for successful classification of cells into cell types, the problem of classification in the brain is more complex than partitioning cells into clusters based on the proximity of their gene expression signatures. The vertebrate nervous system is a uniquely intricate biological network coordinating and controlling all aspects of our behavior. Hence, the neurons that get partitioned into cell types don't exist in isolation. Instead, each neuron in the brain might be connected to as many as hundreds or thousands of other neurons forming neural circuits. Arguably, it is these circuits that act as the fundamental functional units of the brain that generate animal behavior (Figure 1.1). Furthermore, neuronal types and gene expression states in the brain might also be more dynamic in nature. For instance, acquisition of new behaviors or behavioral states might lead to changes in cell type specific gene expression patterns and/or circuit architecture.

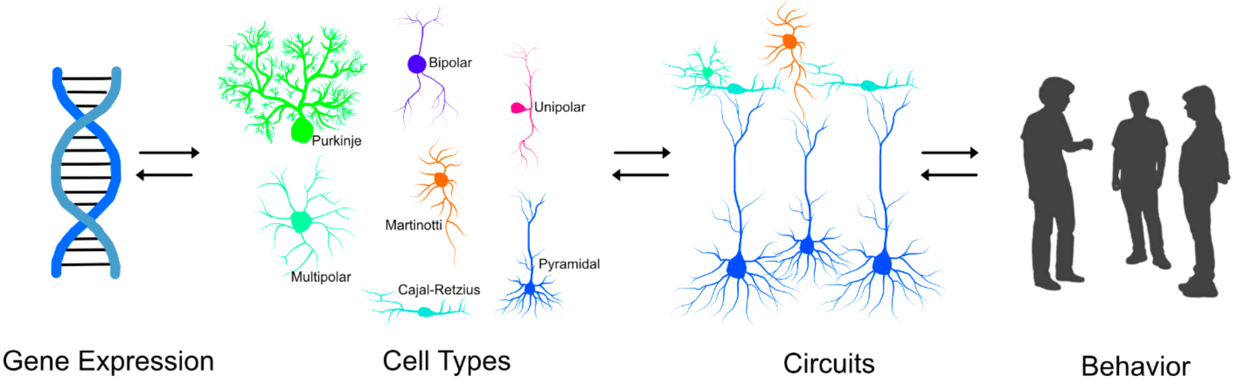


Figure 1.1. The Multiple Scales of the Neuron Identity Problem.

In general, the number of cell types present within a sub-region of the brain might reflect the computational complexity of its functions. In sensory and peripheral nervous system, the characterization of specific cell types as well as their contribution to different behaviors has been studied in great detail (Arber, 2012) . In most of these cases, the reasons for specialization of cell types are also quite apparent. Classic examples include specific motor neurons with distinct molecular, anatomical and electrophysiological properties that make up motor circuits that generate fixed action patterns (Schiff N., 1999). While this principle might also apply to complex circuits of the brain, the number of identifiable cell types present in these circuits might be quite large. For instance, it is believed that there exists 100-150 cell types in the mammalian retina (Zeng and Sanes, 2017). While the function of each of these cell types have not been fully elucidated, there are also tantalizing examples of genetically distinct cell types performing distinct functional roles. For instance, the studies of the JamB retinal ganglion cells illustrate how the anatomy, physiology, molecular properties, spatial localization and receptive fields of these cells are all tailored to their ability to perceive upward motion (Kim et al., 2008).

While the idea that there exists a correspondence between functional properties of circuits and the molecular properties of its constituent cell types is appealing, different functional elements of a circuit may also be distributed across multiple cell types. This is particularly true as we move from peripheral and sensory circuits to central nervous system circuits. In such contexts, it might not be apparent how these circuits might utilize different cell types to mediate different aspects of behavior. For instance, the cerebellar cortex can execute complex motor learning tasks with a handful of molecularly distinct cell types- a classic example of the circuit is greater than the sum of its parts (Llinas and Welsh, 1993).

Therefore, the differing complexities of the brain in different regions necessitates a cell type discovery pipeline that has the ability to parse out fine molecular domains and discover rare molecular states from a large group of neurons. In addition, methodologies are needed to connect these cell types to specific functional roles within their native circuits.

1.3 Maintenance of Cell Type Identity

While neurons might attain their identity after passing through final stages of differentiation, they must remain dynamically responsive to growth factors or neuronal activity while still maintaining their molecular ground state. For instance, a set of activity regulated gene expression programs are known to be initiated in response to changes in intracellular calcium responses (Cohen and Greenberg, 2008). These programs may ultimately alter synaptic and dendritic properties of neurons (West and Greenberg, 2011). While these programs are important for shaping the circuit properties including synaptic strengths during development, they are also important as the animal continues to learn during its adult life. From a behavioral standpoint, the acquisition of new behaviors may also be accompanied by such changes in molecular properties of neurons. For instance, in zebrafish a number of behaviors including shoaling, operant and classical conditioning emerge later in life. What are the changes in cell type or circuit architecture that may lead to these behavioral changes? It is important to consider these dynamic features of neurons in the context of defining cell types in the brain. In other words, we need a contextual classification scheme that recognizes the dynamic and interconnected nature of these specific cell types.

1.4 Cell Type Maps for Investigation into Circuit Function and Disorders of the Brain

An important goal of classifying neurons into specific subtypes is to enable genetic access to specific cell types. This genetic access can allow us to mark and manipulate specific subtypes within a circuit so as to understand how different components of a circuit might contribute to normal animal behavior (Huang and Zeng, 2013; Luo et al., 2018). The discovery of cell type specific genes can serve as fundamental entry points to how the brain generates behaviors in various contexts.

Similarly, cell type mapping efforts are fundamental to understanding brain disorders. The massive diversity of cell types in the mammalian brain has also presented unique challenges for the understanding of dysfunction associated with neuropsychiatric disorders. Because cell atlases provide detailed maps of cells and their roles in tissues, they will also enable us to understand the mechanisms underlying neuropsychiatric disorders at both cell and cellular-ecosystem level. For example, how does a specific neuropathology in the brain arise from mutations in genes that are broadly expressed? Why does there exist an extreme genetic complexity/variability in disorders such as schizophrenia or autism spectrum disorder? An atlas of the brain will help map the cell type of action for genes associated with neurological disorders such as autism, schizophrenia and depression. Ultimately, this will help us understand the unique vulnerabilities conferred to specific cell types as a result of their specific biochemistries. In addition, the detailed molecular profiling of the affected cell types during the development and progression of disease is an important step in understanding the phenotypic consequences of genetic mutations on specific cell types in the brain. Together, these types of knowledge can guide disease modification by selecting for specific gene signatures to pursue in drug screens to fix the desired cellular phenotypes.

1.5 Criteria for defining cell types

There is considerable debate among neuroscientists in the field over the strict definition of a cell type. In principle, it seems obvious that individual neurons should be viewed as belonging to the same type if they serve the same function that differs from functions of other types of neurons. However, defining the function of a neuron at the level of a single cell is a difficult endeavor. The same neuron may perform different functions under different contexts or the function of a neuron may only emerge at the circuit level. Therefore, neuroscientists have relied on other properties to define cell types given that said properties are homogeneous within the population and differ from neurons of other types. The next question that arises then is the following: what are the properties of neurons that should be considered in defining cell types? There are three main categories of properties that are widely used to define cell types (Bota and Swanson, 2007; Fishell and Heintz, 2013; Petilla Interneuron Nomenclature et al., 2008): 1] morphological (Ramón y Cajal, 1899), 2] electrophysiological (Markram et al., 2004) and 3] molecular properties (Cauli et al., 1997) (Figure 1.2). Morphological features may include such properties as dendritic and axonal shapes, branching patterns, soma size and shape. Electrophysiological features may include resting potential, biophysical properties and firing rate. Molecular features may include mRNA or protein composition assayed by in situ hybridization or RNAseq and immunohistochemistry respectively. Two additional category of features 1) connectivity (Shepherd et al., 2019) and 2) developmental or epigenetic states (Kepecs and Fishell, 2014; Yuste, 2005) are equally important (Figure 1.2) but are less frequently incorporated into current studies due to technological challenges. Ideally, a unifying definition of cell types should involve all of these properties. Although such a satisfying correspondence of various features has been observed in some regions of the brain like the retina (Kim et al., 2008; Shekhar et al., 2016; Tsukamoto and Omi, 2017), it is unclear if this is true in many other parts of the brain. There is also the additional need to identify homologous cell types across different species (Arendt et al., 2016).

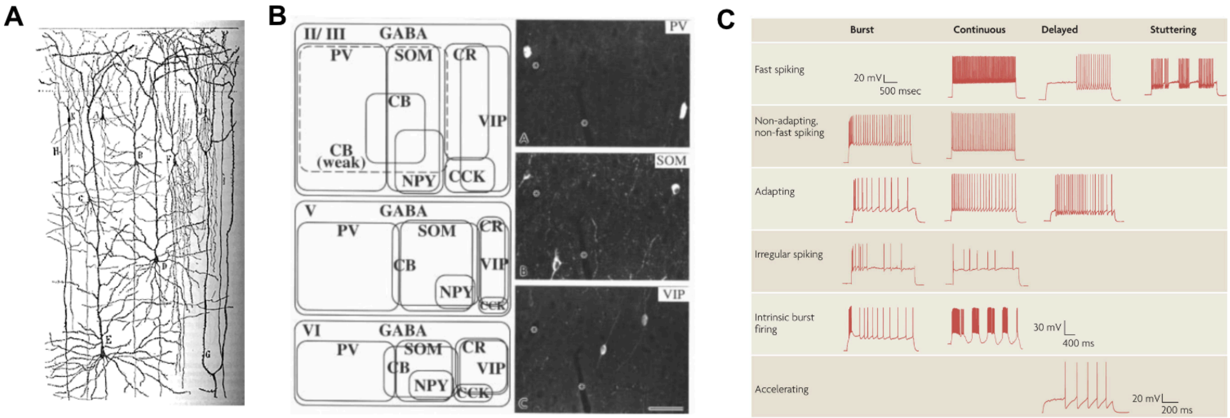


Figure 1.2. Criteria for Cell Type Definition.

Figure 2.1 (Continued). (A) Morphological classification of neurons by Ramón y Cajal (Ramon y Cajal, 1899). The main cell types of the cerebral cortex i.e. small and large pyramidal neurons (A, B, C, D, E) and interneurons (F, K) cells are labeled in the diagram.
(B) Classification of GABAergic neurons based combinatorial marker gene expression from Kawaguchi and Kubota (Kawaguchi, 1997).
(C) Classification of cortical neurons based on intrinsic firing properties from the Petilla Convention (Petilla Interneuron Nomenclature et al., 2008).

A classification scheme based on transcriptomics is advantageous for many reasons. First, high throughput transcriptomics can allow large scale, systematic, quantitative and comprehensive analysis of large regions of the brain. Second, measuring molecular properties over the developmental trajectory of each neuron may offer predictive power to encode other cellular phenotypes such as morphology. Third, molecular definition of cell types enables the identification of cell type markers that can be used to create genetic tools that allow targeting, labeling and manipulation of said types to understand their function (Gerfen et al., 2013; He et al., 2016). Fourth, the nature of transcriptomic data may also allow alignment of cell types across different species using various machine learning techniques (Peng et al., 2019). Last, transcriptomic data can also provide information on potential links between disease associated genes and the cellular locus of action.

1.6 Integrative Neurobiology with Zebrafish

As discussed above, the problem of cell type classification in the brain needs to span multiple scales from measuring gene expression in individual cells to system-level behavioral characterization that seeks to connect molecular features to behavioral outputs. Several features make zebrafish an ideal model organism for tackling this problem.

1) The size of the larval brain is 100,000 neurons, making comprehensive profiling of brain regions and discovery of rare cell types more feasible compared to the mammalian brain.
2) The larval zebrafish brain is transparent, making it possible to interface cell type maps with brain-wide activity maps obtained from calcium imaging.
3) Larval zebrafish also display a range of simple to complex behaviors which can be measured in high-throughput fashion, allowing the behavior profiling of mutants of disease associated genes.

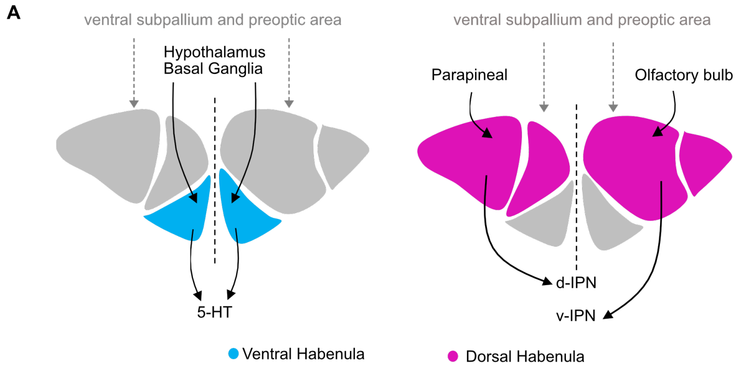
In this thesis, I will mainly focus on a region of the fish brain called the habenula, where I employed high-throughput approaches such as single-cell RNAseq to delineate cell types (Chapter 2). I will present results that follow up on behavioral and brain activity studies of mutant genes to understand the role of cell-type specific genes in various behaviors (Chapter 3). Later, I will also present a vignette on the dissection of cell type specific effects of loss of function mutations of schizophrenia associated genes in zebrafish (Chapter 4). In the following sections, I will introduce the habenula and the variety of functional roles that it plays in generating behavior in both healthy and diseased contexts.

1.7 Introduction to the Habenula

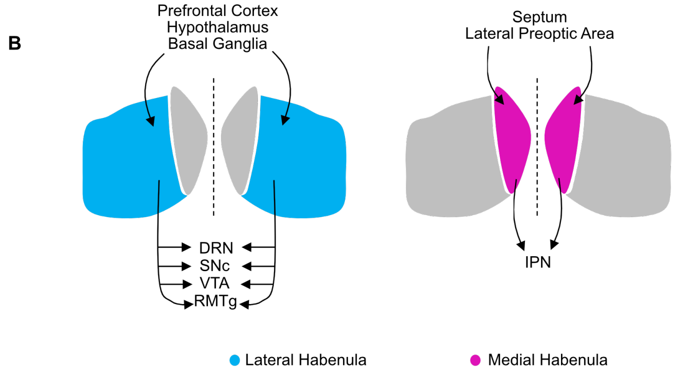
The habenula is a conserved brain structure that is present in all vertebrate brains. First identified by Serres in 1824 and later described in more anatomical detail by Ramon y Cajal, the habenula is a pair of small nuclei located the epithalamus, dorsal to the thalamus (Antolin-Fontes et al., 2015). It serves as a neuroanatomical hub that connects the limbic forebrain to midbrain monoaminergic centers. As such a hub, the habenula plays fundamental roles in vertebrate neurophysiology and behavior (Bianco and Wilson, 2009). It influences a wide range of behaviors, including sleep, pain processing, reward learning, stress and fear (Bianco and Wilson, 2009; Hikosaka, 2010; Namboodiri et al., 2016). Its pathophysiology has also been implicated in neurological disorders such as depression, schizophrenia and addiction (Proulx et al., 2014).

1.7.1 Neural Network of the Habenular Nuclei

As outlined in Figure 1.3, the habenula serves as a neuroanatomical hub between forebrain regions (prefrontal cortex, ventral subpallium, preoptic area, basal ganglia and hypothalamus) and midbrain regions (interpeduncular nucleus, dopaminergic and serotonergic centers). There is also a remarkable conservation of the afferent and efferent connections of the habenula between fish and mammals. With regards to projection patterns, the dorsal habenula in fish were found to be homologous to the medial habenula of mammals, while the ventral habenula in fish displayed homology to the mammalian lateral habenula.



Zebrafish Habenula



Mammalian Habenula

Figure 1.3. Neural Network of the A) Zebrafish and B) Mammalian Habenula.
Abbreviations: 5-HT: serotonin; d-IPN: dorsal Interpeduncular Nucleus; v-IPN: ventral Interpeduncular Nucleus; DRN: Dorsal Raphe Nucleus; SNc: Substantia Nigra; VTA: Ventral Tegmental Area; RMTg: Rostromedial Tegmental Nucleus.

The output projections of the habenula are called the fasciculous retroflexus (fr) in both mammals and fish and are described to go to only to these specific circuits suggesting that they are part of dedicated circuit functions. An important distinction between fish and mammals lie in the L-R asymmetric inputs that go into the dorsal habenula of fish. These represent the input from the parapineal that is known to carry visual information into the left dorsal habenula and input from the olfactory bulb that carries olfactory information specifically to the right habenula.

1.7.2 Varied Functional Roles of the Habenula

A. The Role of Habenula in promoting sleep

Several lines of evidence have directly or indirectly suggested the role of habenula in the regulation of sleep. First, in rats, lesioning the fasciculus retroflexus, the primary output of the habenula leads to reduction of atonia during Rapid Eye Movement (REM) and the amount of time spent in REM sleep (Haun et al., 1992). These aberrant REM sleep phenotypes can be restored by transplanting habenular tissue back into rats. Second, both in slice preparations and in vivo, habenular neurons show circadian rhythmicity (Guilding and Piggins, 2007). Third, medial habenular neurons produce the cytokine interleukin (IL)-18 which is known to promote sleep when injected intracerebroventricularly (Kubota et al., 2001). Fourth, the neural activity of both lateral and medial habenula has been observed to increase significantly during general anesthesia (Abulafia et al., 2009; Herkenham, 1981). These lines of evidence suggest that the habenula plays an important regulatory role in sleep. This regulatory role is particularly interesting when put in context of the co-evolution of the habenula and pineal gland. The pineal gland is well known to regulate circadian rhythm by releasing melatonin and has connections with the habenula (Brainard et al., 2001; Semm et al., 1981). In fish, in particular, photosensitive cells in the parapineal gland relay photic information into the left habenula (Dreosti et al., 2014). Similarly, in mammals, melanopsin-expressing retinal ganglion cells send input to the lateral habenula in addition to the suprachiasmatic nucleus (Hattar et al., 2006). It is therefore not unlikely that the habenula may use this photic information to regulate sleep. However, till date, there has been no exploration of the role of habenula in regulating sleep in zebrafish.

B. The Role of Lateral Habenula (LHb) in Reward-based decision-making and Avoidance of Punishment

As outlined in the neural network above (Figure 1.3), the LHb projects directly and indirectly to monoaminergic centers in the brain. One of the major targets of the LHb is the midbrain dopaminergic neurons which are known to be key regulators of action selection with the goal of maximizing rewards and minimizing punishment.

During action selection to a sensory cue, dopamine neurons are well known to compute the discrepancy between the received and predicted outcome, a variable known as reward prediction error. Typically, dopamine neurons are excited by larger-than-expected rewards and inhibited by smaller-than-expected rewards. It is changes in these dopamine neuron dynamics over time that drive reinforcement learning. Recent studies have revealed that dopamine neurons receive these reward related signals from the LHb that is situated upstream in this neural network. In vivo electrophysiological studies have demonstrated that the LHb neurons are excited by smaller-than-expected rewards whereas they are inhibited by larger-than-expected rewards, a firing pattern that is in direct contrast with the downstream dopamine neurons. Therefore, lateral habenular neurons contribute to reward-based reinforcement learning through inhibitory action on the midbrain dopamine neurons.

Similarly, several lines of evidence suggest that the habenula may have important roles in generating behavioral responses to aversive stimuli, including pain. For instance, in a pavlovian task conducted in monkeys, lateral habenular neurons are strongly activated by those conditioned stimuli that indicated lack of reward or aversive stimuli. Similarly, rats with habenular lesions display impairments in avoidance learning. These results indicate that the LHb, which is often also termed the anti-reward center of the brain, responds strongly to negative value of a stimulus and selects for actions that either lead to reward or avoidance of punishment. Similar studies have been performed in fish which demonstrate the role of the homologous ventral habenula in encoding aversive expectation value and mediating avoidance of danger (Amo et al., 2014).

C. Habenula in Behavioral Responses to Stress

Stimuli that induce stress such as repetition of aversive stimuli, physical constraint, open field exposure and social defeat all lead to activation of the LHb in mammals (Shumake and Gonzalez-Lima, 2003). One of the typical behavioral responses to stress is a suppression of motor activity (Seligman, 1972). This suppression of motor activity is thought to be mediated via an increase in activity in the lateral habenular neurons which in turn strongly inhibit dopaminergic neurons in the substantia nigra or ventral tegmental area (Matsumoto and Hikosaka, 2009). Interestingly, the habenula also seems to integrate the levels of stress received over a long period of time. For instance, in monkeys that have been stressed repeatedly and for a long period of time, the baseline neuronal activity level in the LHb is continuously elevated, leading to a substantial reduction in baseline activity of dopamine neurons (Matsumoto and Hikosaka, 2009). Similarly, congenitally stress-vulnerable rats that are exposed to repeated aversive stimuli display helpless behavior that is accompanied by a higher metabolic activity in the LHb and lower activity in the dopaminergic neurons in the VTA (Shumake et al., 2003). It is therefore thought that the stress induced activation of LHb leads to strong inhibition of dopamine neurons which subsequently leads to suppression of motor activity.

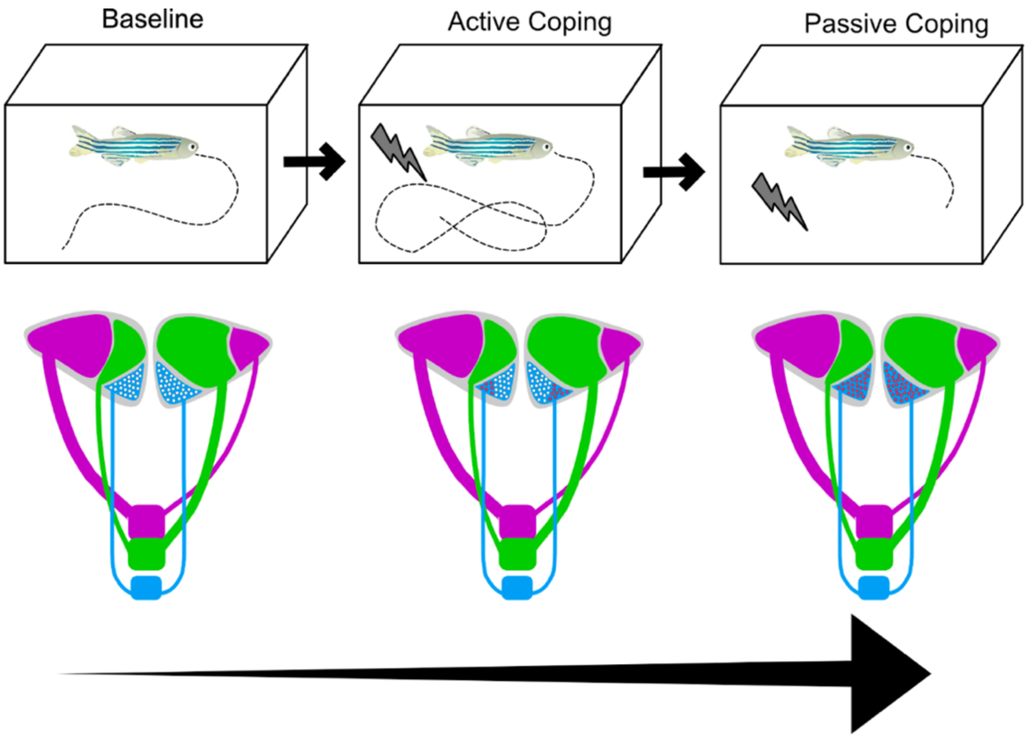


Figure 1.4. Role of Ventral Habenula in Mediating a Behavioral Switch from Active to Passive Coping during inescapable stressors. [Adapted from: (Rodriguez-Sosa et al., 2019)]

A recent study in fish also indicated that the habenula has an important role in generating behavioral strategies to strong inescapable stressors and specifically in mediating a behavioral switch from active to passive coping (Figure 1.4) (Andalman et al., 2019).

D. Habenula in Fear and Anxiety

The medial habenula regulates the expression of fear in zebrafish. Following silencing with tetanus toxin or lesioning with nitroreductase (Agetsuma et al., 2010; Lee et al., 2010), enhanced freezing was seen in aversive conditioning paradigms in fish. The freezing response in habenula-lesioned fish was experience-dependent. On the other hand, the left medial habenula in fish has been found to attenuate fear responses in larval zebrafish. In similar studies in mice, the septum-medial habenula pathway was also found to regulate anxiety and fear (Yamaguchi et al., 2013), suggesting that this conserved circuit has dedicated functional roles in fear in many vertebrate animals.

E. Habenular Dysfunction in Psychiatric Disorders

An increasing number of studies have implicated the habenula in psychiatric disorders associated with dysregulated reward circuitry function, most notably mood disorders such as depression, anxiety and others such as schizophrenia and addiction (Li et al., 2019; Schafer et al., 2018; Shumake and Gonzalez-Lima, 2003; Velasquez et al., 2014; Yang et al., 2018b).

Prolonged exposure to pain and stress in humans has been directly associated with the development of major depression (Seligman, 1972). Several studies have directly demonstrated that hyperactivation of lateral habenula is associated with depressive phenotypes (Cui et al., 2018). For instance, in rat models of depression, neural activity in the lateral habenula is markedly increased. Similarly, local infusion of ketamine, NMDAR antagonist in the lateral habenula rapidly alleviated depression like symptoms in rats (Yang et al., 2018a). This effect on depression could be mediated by the removal of the inhibitory effect of LHb on dopamine neurons. Similarly, a major theory of depression also implicates altered serotonin levels in the pathophysiology of the depression. This theory is supported by the action of serotonin reuptake inhibitors on the successful treatment of depression (Middlemiss et al., 2002). It is also possible that the upstream modulation of serotonin in the brain may be a result of increased or decreased inhibitory signal from the LHb neurons. In lines with these direct and indirect evidence for the association of habenula and depression, successful treatment of depression patients using deep brain stimulation of the habenula have been reported (Sartorius and Henn, 2007).

Following from its major role in reward processing, habenular circuitry is also important in the development of addiction (Velasquez et al., 2014). Neurons in the lateral habenula are strongly implicated in alcohol and cocaine abuse. During cocaine withdrawal, activity in the lateral habenula is elevated due to a reduction in release of GABA from the upstream entopeduncular nucleus leading to a general increase in glutamatergic drive (Meye et al., 2016). Furthermore, activating the lateral habenula reduces voluntary ethanol consumption in rats (Li et al., 2016). Similarly, although nicotinic acetylcholine receptors are expressed in many regions of the brain, the neuronal groups that mediate nicotine withdrawal responses are concentrated in the medial habenula (Salas et al., 2009). Genome wide association studies of tobacco addiction have produced highly consistent results showing the role of variants in the alpha3, alpha5 and beta4 nAChR gene clusters in tobacco use and addiction (Saccone et al., 2010). Brain expression of all of these receptors is particularly high in the medial (dorsal) habenula. In fact, the medial habenula is the only region that co-expresses the mRNA of all of these subunits at very high and moderate levels (Baldwin et al., 2011).

A rich molecular profile of neurons within the habenula will ultimately also help us understand the molecular determinants of selective vulnerability of the habenula to these neurological disorders.

1.7.3 Molecular Anatomical and Functional Asymmetry in the Teleostan Habenula

A striking feature of the teleostan habenula is that it displays a remarkable left-right asymmetry in molecular, anatomical and functional properties. By larval stages, several molecular and neuroanatomical asymmetries distinguish the left and right sides of the habenula. First, the parapineal which is a light sensitive organ present in many teleosts is located on the left side of the habenula and innervates only the left habenula. This asymmetric migration of the parapineal is the first sign of asymmetry within the habenula and is known to be mediated by left-sided nodal signaling leading to asymmetric neurogenesis across the two sides of the habenula (Figure 1.5) (Concha et al., 2000; Roussigne et al., 2009). This leads to differences in the proportion of neuronal subtypes with distinct gene expression between the left and right habenula. A large proportion of the neurons in the left habenula adopt a dorso-lateral identity and are defined by the expression of *nptx2a*. However, a large proportion of neurons in the right habenula adopt a dorso-medial identity and are defined by the expression of *gpr151*. This is also followed by the asymmetric expression of Neuropilin1a (*nrp1a*), a receptor for class III Semphorins (*Sema3D*) in the left habenula. Together, Sema3D and nrp1a act in concert to guide neurons on the left habenula to innervate downstream targets differently than those on the right side (Kuan et al., 2007) (Figure 1.5).

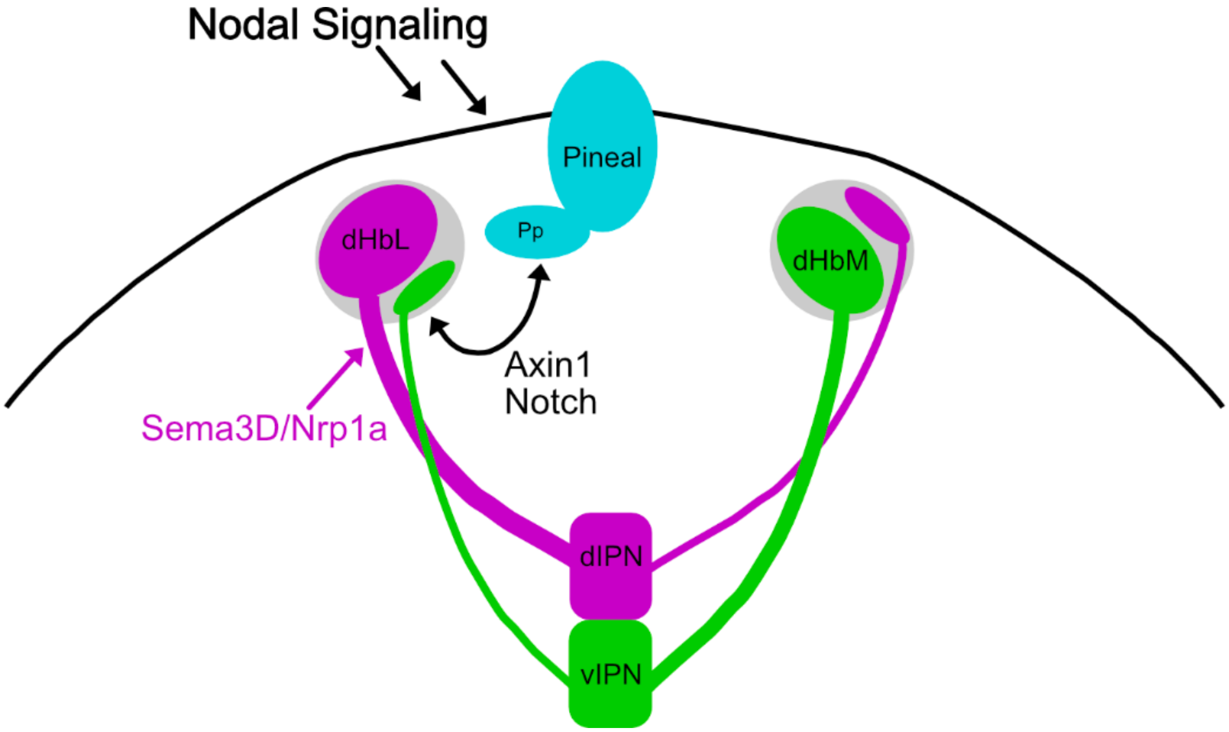


Figure 1.5. Generation of Asymmetry in the Dorsal Habenula via asymmetric Nodal Signaling.

This leads to asymmetry in the out-projections from the two lobes of the habenula (Figure 1.6). A large portion of the left habenula projects to the dorsal Interpeduncular Nucleus (dIPN) whereas the right habenula projects to the ventral IPN (vIPN). These circuitries go to different targets downstream of the IPN. The dIPN neurons pass through the dorsal raphe to reach the dorsal tegmental area (DTA), a putative region corresponding to the mammalian periaqueductral grey (PAG) which regulates fight, flight and freezing behaviors (George et al., 2019). The vIPN neurons on the other hand project to the median raphe (MR). These distinct pathways therefore have been shown to mediate distinct behaviors. For instance, in social conflict resolution, the dHbL-dIPN-DTA circuit promotes winning behaviors whereas the dHbM-vIPN-MR circuit promotes losing behaviors (Chou et al., 2016).

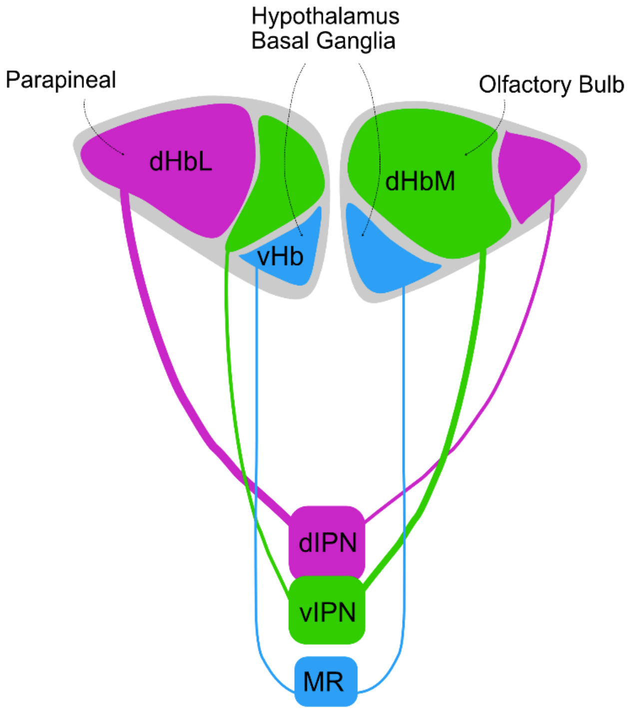


Figure 1.6. Input and Output Projections Patterns of the Zebrafish Habenula.

Similar to the output projections, there also exist distinct asymmetries in the input projections of the left and right dorsal habenula (Figure 1.5). The left dorsal habenula receives input from the light sensitive parapineal whereas the right dorsal habenula receives input from the olfactory bulb. Therefore, sensory responses to light and odor are lateralized in the dorsal habenula (Dreosti et al., 2014). These light and odor responsive neurons have important roles in mediating light preference (Zhang et al., 2017) and attraction to specific concentrations of different odor cues (Krishnan et al., 2014) (Figure 1.7).

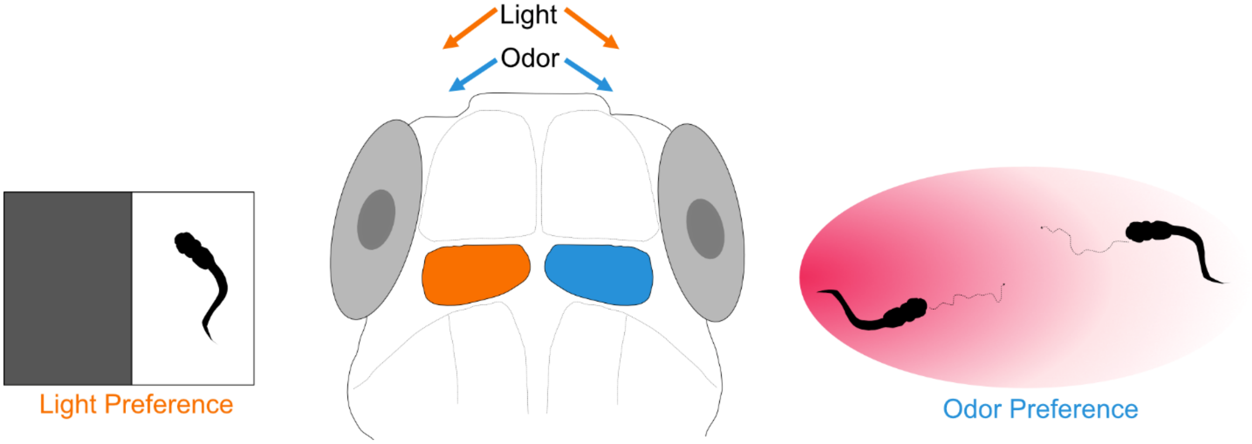


Figure 1.7. Functional Asymmetry in the Zebrafish Habenula.

In addition to the described asymmetry in functional properties and connectivity, a number of L-R asymmetric genes such as *adcyap1a, nptx2a, tac1, hacta* have been described in the habenula by RNA-FISH (deCarvalho et al., 2014). However, it is unclear if these domains with distinct gene expression patterns represent multiple neuronal subtypes. Furthermore, the genes (*nptx2a*, *gpr151* and *aoc1*) that are typically used to distinguish the three subnuclei of the zebrafish habenula: dorso-lateral, dorso-medial and ventral subnuclei are broadly expressed. For this reason, these sub-region-specific genes are often used as genetic handles in functional studies (Agetsuma et al., 2010; Amo et al., 2014; Chou et al., 2016). However, it is unclear if these broader sub-regions may harbor multiple neuronal types.

1.8 Aims and Approaches

In 2014, at the beginning of my graduate work in the Schier lab, low throughput single-cell RNAseq techniques such as Smart-seq2 were being actively used to delineate cell types in various systems (Ramskold et al., 2012; Shalek et al., 2013) including blastomeres in early zebrafish development (Satija et al., 2015). These technologies were being actively translated to various regions of the mammalian brain. I began my graduate work extending scRNA-seq to the zebrafish brain, which presented a host of technical challenges due to the small size of the neurons. I optimized and generated a robust workflow for dissection, dissociation and generation of single cell libraries from the zebrafish brain using both fluorescence activated cell sorting and unbiased dissociation. We used these approaches and multiple scRNA-seq modalities to generate a comprehensive molecular atlas of the zebrafish habenula (Chapter 2).

Various features made the zebrafish habenula an ideal system to interrogate the problem of neuron identity in a comprehensive manner, most notably the following:

1) The zebrafish habenula consists of ~1000 neurons in larva stages allowing a complete profiling of all cells.
2) Its projection patterns and functional repertoire is also well described.
3) There are distinct functional and molecular left-right asymmetry built into the structure making it an ideal candidate region for interfacing molecular and functional maps (see Section on Functional Asymmetry in the Zebrafish Habenula).

Next, we utilized the single cell atlas of the habenula to overlay functional information onto the cell types generated by single cell RNAseq using two approaches (Chapter 3):

1) Spatial mapping of *cfos* responses onto an *in situ* atlas generated by utilizing markers from the single-cell RNAseq study.
2) Generation of CRISPR knockouts of restricted marker genes and subsequent characterization of brain activity and behaviors in these mutant animals.

Last, we utilized single-cell RNAseq to understand the molecular underpinnings of behavioral and brain activity phenotypes in the forebrain using mutants of schizophrenia associated genes (Chapter 4), generating testable hypothesis for the function of these genes. We identified cell type specific vulnerabilities to the loss of these genes. Together, these studies provide a scaffold for the unbiased discovery of cell types from various brain regions as well as the discovery of cell types underlying specific functional phenotypes in the brain.

REFERENCES

Abulafia, R., Zalkind, V., and Devor, M. (2009). Cerebral activity during the anesthesia-like state induced by mesopontine microinjection of pentobarbital. J Neurosci *29*, 7053-7064.

Agetsuma, M., Aizawa, H., Aoki, T., Nakayama, R., Takahoko, M., Goto, M., Sassa, T., Amo, R., Shiraki, T., Kawakami, K*., et al.* (2010). The habenula is crucial for experience-dependent modification of fear responses in zebrafish. Nat Neurosci *13*, 1354-1356.

Amo, R., Fredes, F., Kinoshita, M., Aoki, R., Aizawa, H., Agetsuma, M., Aoki, T., Shiraki, T., Kakinuma, H., Matsuda, M*., et al.* (2014). The habenulo-raphe serotonergic circuit encodes an aversive expectation value essential for adaptive active avoidance of danger. Neuron *84*, 1034-1048.

Andalman, A.S., Burns, V.M., Lovett-Barron, M., Broxton, M., Poole, B., Yang, S.J., Grosenick, L., Lerner, T.N., Chen, R., Benster, T*., et al.* (2019). Neuronal Dynamics Regulating Brain and Behavioral State Transitions. Cell *177*, 970-985 e920.

Antolin-Fontes, B., Ables, J.L., Gorlich, A., and Ibanez-Tallon, I. (2015). The habenulo-interpeduncular pathway in nicotine aversion and withdrawal. Neuropharmacology *96*, 213-222.

Arber, S. (2012). Motor circuits in action: specification, connectivity, and function. Neuron *74*, 975-989.

Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. Nat Rev Genet *9*, 868-882.

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D*., et al.* (2016). The origin and evolution of cell types. Nat Rev Genet *17*, 744-757.

Baldwin, P.R., Alanis, R., and Salas, R. (2011). The Role of the Habenula in Nicotine Addiction. J Addict Res Ther *S1*.

Bianco, I.H., and Wilson, S.W. (2009). The habenular nuclei: a conserved asymmetric relay station in the vertebrate brain. Philos Trans R Soc Lond B Biol Sci *364*, 1005-1020.

Bota, M., and Swanson, L.W. (2007). The neuron classification problem. Brain Res Rev *56*, 79-88.

Brainard, G.C., Hanifin, J.P., Greeson, J.M., Byrne, B., Glickman, G., Gerner, E., and Rollag, M.D. (2001). Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. J Neurosci *21*, 6405-6412.

Cauli, B., Audinat, E., Lambolez, B., Angulo, M.C., Ropert, N., Tsuzuki, K., Hestrin, S., and Rossier, J. (1997). Molecular and physiological diversity of cortical nonpyramidal cells. J Neurosci *17*, 3894-3906.

Chou, M.Y., Amo, R., Kinoshita, M., Cherng, B.W., Shimazaki, H., Agetsuma, M., Shiraki, T., Aoki, T., Takahoko, M., Yamazaki, M.*, et al.* (2016). Social conflict resolution regulated by two dorsal habenular subregions in zebrafish. Science *352*, 87-90.

Cohen, S., and Greenberg, M.E. (2008). Communication between the synapse and the nucleus in neuronal development, plasticity, and disease. Annu Rev Cell Dev Biol *24*, 183-209.

Concha, M.L., Burdine, R.D., Russell, C., Schier, A.F., and Wilson, S.W. (2000). A nodal signaling pathway regulates the laterality of neuroanatomical asymmetries in the zebrafish forebrain. Neuron *28*, 399-409.

Cui, Y., Yang, Y., Ni, Z., Dong, Y., Cai, G., Foncelle, A., Ma, S., Sang, K., Tang, S., Li, Y.*, et al.* (2018). Astroglial Kir4.1 in the lateral habenula drives neuronal bursts in depression. Nature *554*, 323-327.

deCarvalho, T.N., Subedi, A., Rock, J., Harfe, B.D., Thisse, C., Thisse, B., Halpern, M.E., and Hong, E. (2014). Neurotransmitter map of the asymmetric dorsal habenular nuclei of zebrafish. Genesis *52*, 636-655.

Dreosti, E., Vendrell Llopis, N., Carl, M., Yaksi, E., and Wilson, S.W. (2014). Left-right asymmetry is required for the habenulae to respond to both visual and olfactory stimuli. Curr Biol *24*, 440-445.

Fishell, G., and Heintz, N. (2013). The neuron identity problem: form meets function. Neuron *80*, 602-612.

George, D.T., Ameli, R., and Koob, G.F. (2019). Periaqueductal Gray Sheds Light on Dark Areas of Psychopathology. Trends Neurosci *42*, 349-360.

Gerfen, C.R., Paletzki, R., and Heintz, N. (2013). GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. Neuron *80*, 1368-1383.

Guilding, C., and Piggins, H.D. (2007). Challenging the omnipotence of the suprachiasmatic timekeeper: are circadian oscillators present throughout the mammalian brain? Eur J Neurosci *25*, 3195-3216.

Hattar, S., Kumar, M., Park, A., Tong, P., Tung, J., Yau, K.W., and Berson, D.M. (2006). Central projections of melanopsin-expressing retinal ganglion cells in the mouse. J Comp Neurol *497*, 326-349.

Haun, F., Eckenrode, T.C., and Murray, M. (1992). Habenula and thalamus cell transplants restore normal sleep behaviors disrupted by denervation of the interpeduncular nucleus. J Neurosci *12*, 3282-3290.

He, M., Tucciarone, J., Lee, S., Nigro, M.J., Kim, Y., Levine, J.M., Kelly, S.M., Krugikov, I., Wu, P., Chen, Y*., et al.* (2016). Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex. Neuron *92*, 555.

Herkenham, M. (1981). Anesthetics and the habenulo-interpeduncular system: selective sparing of metabolic activity. Brain Res *210*, 461-466.

Hikosaka, O. (2010). The habenula: from stress evasion to value-based decision-making. Nat Rev Neurosci *11*, 503-513.

Huang, Z.J., and Zeng, H. (2013). Genetic approaches to neural circuits in the mouse. Annu Rev Neurosci *36*, 183-215.

Kawaguchi, Y.K., Y. (1997). GABAergic cell subtypes and their synaptic connections in rat frontal cortex. Cereb Cortex, 476-486

Kepecs, A., and Fishell, G. (2014). Interneuron cell types are fit to function. Nature *505*, 318-326.

Kim, I.J., Zhang, Y., Yamagata, M., Meister, M., and Sanes, J.R. (2008). Molecular identification of a retinal cell type that responds to upward motion. Nature *452*, 478-482.

Krishnan, S., Mathuru, A.S., Kibat, C., Rahman, M., Lupton, C.E., Stewart, J., Claridge-Chang, A., Yen, S.C., and Jesuthasan, S. (2014). The right dorsal habenula limits attraction to an odor in zebrafish. Curr Biol *24*, 1167-1175.

Kuan, Y.S., Yu, H.H., Moens, C.B., and Halpern, M.E. (2007). Neuropilin asymmetry mediates a left-right difference in habenular connectivity. Development *134*, 857-865.

Kubota, T., Fang, J., Brown, R.A., and Krueger, J.M. (2001). Interleukin-18 promotes sleep in rabbits and rats. Am J Physiol Regul Integr Comp Physiol *281*, R828-838.

Lee, A., Mathuru, A.S., Teh, C., Kibat, C., Korzh, V., Penney, T.B., and Jesuthasan, S. (2010). The habenula prevents helpless behavior in larval zebrafish. Curr Biol *20*, 2211-2216.

Li, J., Yang, S., Liu, X., Han, Y., Li, Y., Feng, J., and Zhao, H. (2019). Hypoactivity of the lateral habenula contributes to negative symptoms and cognitive dysfunction of schizophrenia in rats. Exp Neurol *318*, 165-173.

Li, J., Zuo, W., Fu, R., Xie, G., Kaur, A., Bekker, A., and Ye, J.H. (2016). High Frequency Electrical Stimulation of Lateral Habenula Reduces Voluntary Ethanol Consumption in Rats. Int J Neuropsychopharmacol.

Llinas, R., and Welsh, J.P. (1993). On the cerebellum and motor learning. Curr Opin Neurobiol *3*, 958-965.

Luo, L., Callaway, E.M., and Svoboda, K. (2008). Genetic dissection of neural circuits. Neuron *57*, 634-660.

Luo, L., Callaway, E.M., and Svoboda, K. (2018). Genetic Dissection of Neural Circuits: A Decade of Progress. Neuron *98*, 256-281.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. Nat Rev Neurosci *5*, 793-807.

Masland, R.H. (2004). Neuronal cell types. Curr Biol *14*, R497-500.

Matsumoto, M., and Hikosaka, O. (2009). Representation of negative motivational value in the primate lateral habenula. Nat Neurosci *12*, 77-84.

Meye, F.J., Soiza-Reilly, M., Smit, T., Diana, M.A., Schwarz, M.K., and Mameli, M. (2016). Shifted pallidal co-release of GABA and glutamate in habenula drives cocaine withdrawal and relapse. Nat Neurosci *19*, 1019-1024.

Middlemiss, D.N., Price, G.W., and Watson, J.M. (2002). Serotonergic targets in depression. Curr Opin Pharmacol *2*, 18-22.

Namboodiri, V.M., Rodriguez-Romaguera, J., and Stuber, G.D. (2016). The habenula. Curr Biol *26*, R873-R877.

Peng, Y.R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G.S., van Zyl, T., Do, M.T.H., Regev, A., and Sanes, J.R. (2019). Molecular Classification and Comparative Taxonomics of Foveal and Peripheral Cells in Primate Retina. Cell *176*, 1222-1237 e1222.

Petilla Interneuron Nomenclature, G., Ascoli, G.A., Alonso-Nanclares, L., Anderson, S.A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., Buzsaki, G., Cauli, B., Defelipe, J.*, et al.* (2008). Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. Nat Rev Neurosci *9*, 557-568.

Proulx, C.D., Hikosaka, O., and Malinow, R. (2014). Reward processing by the lateral habenula in normal and depressive behaviors. Nat Neurosci *17*, 1146-1152.

Ramón y Cajal, S. (1899). La Textura del Sistema Nerviosa del Hombre y los Vertebrados,. (Moya (Primera Edicion), Madrid.

Ramon y Cajal, S.S., N., and Swanson, L.W. (trans.) (1899). Histology of the Nervous System (Oxford University Press).

Ramskold, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C.*, et al.* (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol *30*, 777-782.

Rodriguez-Sosa, N., Biswas, S., and Shabel, S.J. (2019). Coping with Stress, One Habenula Neuron at a Time. Neuron *102*, 520-522.

Roussigne, M., Bianco, I.H., Wilson, S.W., and Blader, P. (2009). Nodal signalling imposes left-right asymmetry upon neurogenesis in the habenular nuclei. Development *136*, 1549-1557.

Saccone, N.L., Schwantes-An, T.H., Wang, J.C., Grucza, R.A., Breslau, N., Hatsukami, D., Johnson, E.O., Rice, J.P., Goate, A.M., and Bierut, L.J. (2010). Multiple cholinergic nicotinic receptor genes affect nicotine dependence risk in African and European Americans. Genes Brain Behav *9*, 741-750.

Salas, R., Sturm, R., Boulter, J., and De Biasi, M. (2009). Nicotinic receptors in the habenulo-interpeduncular system are necessary for nicotine withdrawal in mice. J Neurosci *29*, 3014-3018.

Sartorius, A., and Henn, F.A. (2007). Deep brain stimulation of the lateral habenula in treatment resistant major depression. Med Hypotheses *69*, 1305-1308.

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

Schafer, M., Kim, J.W., Joseph, J., Xu, J., Frangou, S., and Doucet, G.E. (2018). Imaging Habenula Volume in Schizophrenia and Bipolar Disorder. Front Psychiatry *9*, 456.

Schiff N., U.R., Fred Plum and Rodolfo Llinás (1999). Words without Mind. Journal of Cognitive Neuroscience *11*.

Seligman, M.E. (1972). Learned helplessness. Annu Rev Med *23*, 407-412.

Semm, P., Schneider, T., and Vollrath, L. (1981). Morphological and electrophysiological evidence for habenular influence on the guinea-pig pineal gland. J Neural Transm *50*, 247-266.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., *et al.* (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236-240.

Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M., *et al.* (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell *166*, 1308-1323 e1330.

Shepherd, G.M., Marenco, L., Hines, M.L., Migliore, M., McDougal, R.A., Carnevale, N.T., Newton, A.J.H., Surles-Zeigler, M., and Ascoli, G.A. (2019). Neuron Names: A Gene- and Property-Based Name Format, With Special Reference to Cortical Neurons. Front Neuroanat *13*, 25.

Shumake, J., Edwards, E., and Gonzalez-Lima, F. (2003). Opposite metabolic changes in the habenula and ventral tegmental area of a genetic model of helpless behavior. Brain Res *963*, 274-281.

Shumake, J., and Gonzalez-Lima, F. (2003). Brain systems underlying susceptibility to helplessness and depression. Behav Cogn Neurosci Rev *2*, 198-221.

Tsukamoto, Y., and Omi, N. (2017). Classification of Mouse Retinal Bipolar Cells: Type-Specific Connectivity with Special Reference to Rod-Driven AII Amacrine Pathways. Front Neuroanat *11*, 92.

Velasquez, K.M., Molfese, D.L., and Salas, R. (2014). The role of the habenula in drug addiction. Front Hum Neurosci *8*, 174.

West, A.E., and Greenberg, M.E. (2011). Neuronal activity-regulated gene transcription in synapse development and cognitive function. Cold Spring Harb Perspect Biol *3*.

Yamaguchi, T., Danjo, T., Pastan, I., Hikida, T., and Nakanishi, S. (2013). Distinct roles of segregated transmission of the septo-habenular pathway in anxiety and fear. Neuron *78*, 537-544.

Yang, Y., Cui, Y., Sang, K., Dong, Y., Ni, Z., Ma, S., and Hu, H. (2018a). Ketamine blocks bursting in the lateral habenula to rapidly relieve depression. Nature *554*, 317-322.

Yang, Y., Wang, H., Hu, J., and Hu, H. (2018b). Lateral habenula in the pathophysiology of depression. Curr Opin Neurobiol *48*, 90-96.

Yuste, R. (2005). Origin and classification of neocortical interneurons. Neuron *48*, 524-527.

Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. Nat Rev Neurosci *18*, 530-546.

Zhang, B.B., Yao, Y.Y., Zhang, H.F., Kawakami, K., and Du, J.L. (2017). Left Habenula Mediates Light-Preference Behavior in Zebrafish via an Asymmetrical Visual Pathway. Neuron *93*, 914-928 e914.

Chapter 2: Comprehensive Identification and Spatial Mapping of Habenular Neuronal Types Using Single-Cell RNASeq

2.1 PREFACE

A version of this chapter was published in Current Biology on April 2, 2018 (Pandey et al., 2018). Alexander F. Schier and I conceived the study. I designed the study with input from Alexander F. Schier. I optimized the protocol for generating single cell libraries from the zebrafish brain, collected all experimental data. Karthik Shekhar helped with analysis and interpretation of results. Alexander F. Schier and Aviv Regev supported this project.

2.2 ABSTRACT

The identification of cell types and marker genes is critical for dissecting neural development and function, but the size and complexity of the brain has hindered the comprehensive discovery of cell types. We combined single-cell RNA-seq (scRNA-seq) with anatomical brain registration to create a comprehensive map of the zebrafish habenula, a conserved forebrain hub involved in pain processing and learning. Single-cell transcriptomes of ∼13,000 habenular cells with 4× cellular coverage identified 18 neuronal types and dozens of marker genes. Registration of marker genes onto a reference atlas created a resource for anatomical and functional studies. Strikingly, despite brain growth and functional maturation, cell types were retained between the larval and adult habenula. This study provides a gene expression atlas to dissect

habenular development and function and offers a general framework for the comprehensive characterization of other brain regions.

2.3 INTRODUCTION

2.3.1 Comprehensive classification of cell types

The study of formation and function of neural circuits relies on the ability to identify specific cell types that are defined by location, morphology, connectivity and molecular composition. Classical histological and gene expression analyses have recently been extended to single-cell technologies that enable *de novo* identification of cell types based on their transcriptomes (Chen et al., 2017; Pollen et al., 2015; Pollen et al., 2014; Poulin et al., 2016; Tasic et al., 2016; Zeisel et al., 2015; Zeng and Sanes, 2017). Such studies have provided valuable resources for cataloguing cell types but are limited in their comprehensive classification by the large number and diversity of neurons in vertebrate brains. This complexity results in low sampling of rare cell types even with recent technologies that allow profiling of thousands of individual neurons in a single experiment (Klein et al., 2015; Macosko et al., 2015). With the possible exception of a single class of interneurons in the mouse retina, we lack comprehensive catalogues of cell types in any region of the vertebrate brain (Shekhar et al., 2016). One scenario in which sampling limits can be overcome is in the case of specific and conserved brain regions in animals with compact size. To test this approach, we analyzed the zebrafish habenula, a small forebrain region that is composed of approximately ~1,500 neurons at the larval stage.

Current anatomical and molecular analysis partitions the zebrafish habenula into three major sub-regions: the *nptx2a* expressing dorso-lateral domain, the *gpr151/pou4f1* expressing dorso-medial domain, and the *aoc1* expressing ventral domain. Neurons in these domains project to distinct downstream regions in the interpeduncular nucleus (IPN) and raphe nucleus, thus mediating distinct behavioral outputs (Beretta et al., 2012; Bianco and Wilson, 2009). These domains are also homologous to distinct domains in the mouse habenula(Okamoto and Aizawa, 2013). For instance, the ventral habenula of zebrafish shares gene expression and projection patterns with the mammalian lateral habenula (Amo et al., 2010). Furthermore, domain-specific genes are often used as genetic handles in functional studies (Agetsuma et al., 2010; Amo et al., 2014; Chou et al., 2016).

It has been unclear, however, whether individual neurons in these sub-nuclei represent a single neuronal type or a mixture of multiple types. In addition, the zebrafish habenula displays a remarkable left-right (L-R) asymmetry in gene expression and functionality (Okamoto, 2014). A number of genes such as *adcyap1a, nrp1a, tac1, tac3a, slc5a7a* are left-right asymmetric in the dorsal habenula (Amo et al., 2010; Biran et al., 2012; deCarvalho et al., 2014; Hong et al., 2013; Kuan et al., 2007). Recent studies have also shown left-right asymmetry in functional responses to light and odor in the left and right habenula, respectively (Dreosti et al., 2014; Jetti et al., 2014; Kishimoto et al., 2013). It is also unclear if these neuronal ensembles represent transcriptionally distinct neuronal types. A comprehensive definition of habenular neuronal types is therefore needed to study its development and anatomy, and relate molecularly defined neuronal types to functional roles.

To address this challenge, we combined scRNA-seq with anatomical brain registration and created a gene expression atlas composed of more than a dozen distinct neuronal types. We find that neuronal types are anatomically organized into spatially segregated sub-regions and are stable between larval and adult stages. We show that the reference atlas enables comparison of molecularly defined neuronal types with those defined by neural activity. Our approach constitutes a general framework for future studies aiming to comprehensively characterize other brain regions.

2.4 RESULTS

2.4.1 Isolation and Transcriptional Profiling of Single Larval Zebrafish Neurons

Since scRNA-seq had not been previously applied to zebrafish neurons, we devised and optimized a robust protocol for dissociation and capture of single neurons from the zebrafish brain. We found that successful experiments required gentle trituration, reduced processing time post dissociation (<30 minutes) and minimal sort pressure during fluorescence activated cell sorting (FACS) (Methods). We sorted habenular cells using the *gng8*-GFP transgenic line (Hong et al., 2013), which selectively labels most neurons in the habenula, except for a small ventral subpopulation (deCarvalho et al., 2014). Heterogeneity in this subpopulation was captured using adult animals as described later.

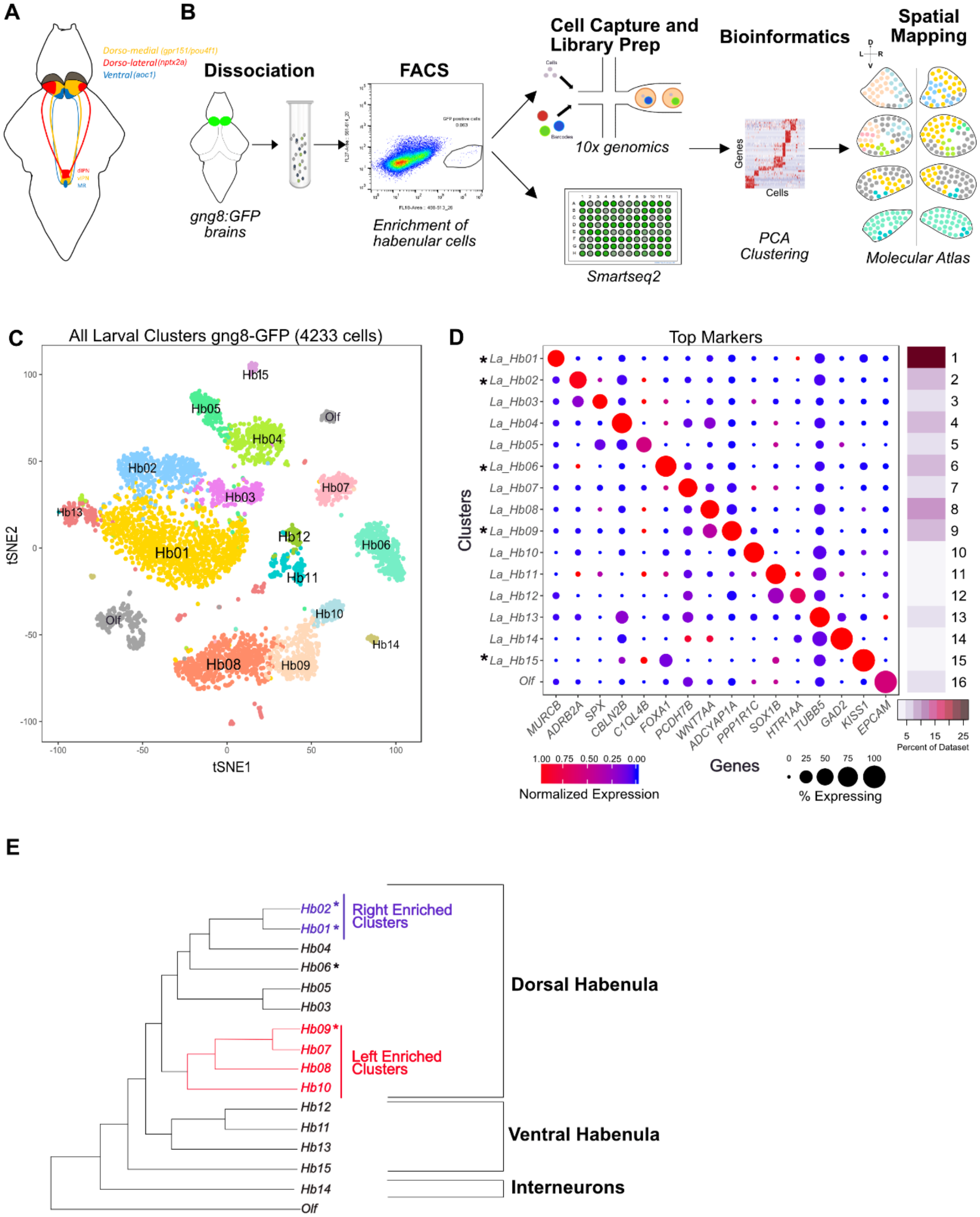


Figure 2.1. Unbiased Clustering of scRNA-seq Data Identifies 15 Molecular Distinct Neuronal Clusters in the Larval Habenula.

(A) Schematic of the zebrafish habenula showing the anatomical subdivisions corresponding to the dorso-medial (orange), dorso-lateral (red) and ventral (blue) regions. These subdivisions are known to have distinct gene expression patterns and functionality.

(B) Overview of the experimental strategy. Transgenic heads with *gng8*-GFP positive cells were dissected, pooled and dissociated, followed by enrichment of GFP$^+$ habenular cells using fluorescent activated cell sorting (FACS). Single cell libraries were prepared using droplet-based droplet and plate-

Figure 2.1 (Continued). based Smart-seq2. Raw reads were processed to obtain a gene expression matrix (genes x cells). PCA and graph clustering was used to divide cells into clusters and identify cluster specific markers. Validation and spatial localization was performed using fluorescent RNA *in situ* hybridization (FISH) of statistically significant cluster-specific markers (see STAR Methods).

(C) 2D visualization of single cell clusters using t-distributed Stochastic Neighbor Embedding (tSNE). Individual points correspond to single cells and are color-coded according to their cluster membership determined by graph-based clustering. The tSNE mapping was only used for post hoc visualization of the clustering but not to define the clusters themselves.

(D) Gene Expression profiles (columns) of select cluster-specific markers identified through differential expression analysis (DEA) of previously known (labeled with an asterisk (*)) and new habenular types (rows). Bar on the right displays the percent of total dataset represented in every cluster, showing the abundance of each cell type found by clustering analysis.

(E) A dendrogram representing global inter-cluster transcriptional relationships. The dendrogram was built by performing hierarchical clustering (correlation distance, average linkage) on the average gene-expression profiles for each cluster restricting to the highly variable genes in the dataset.

We used two complementary scRNA-seq platforms (Figure 2.1B, Methods): (1) Massively-parallel droplet based, 3’ end scRNA-Seq (commercial 10X Chromium platform) (Zheng et al., 2017). (2) Lower-throughput, plate-based, full-length scRNA-Seq (SMART-Seq2) (Picelli et al., 2013). We obtained data from 4,365 larval cells using droplet-based scRNA-seq (henceforth droplet dataset) sequenced at a median depth of 66,263 reads per cell, and 1,152 larval cells using SMART-seq2 (henceforth SS2 dataset), sequenced at a median depth of 1 million reads per cell. Despite the low levels of RNA in zebrafish neurons (~ 6 times less than mouse dendritic cells which are of comparable size; Figures A.1 A and A.1B), we detected on average 1,350 (droplets) and 3,850 (SS2) genes/cell (Figures A.1C, A.4A). Thus, multiple scRNA-seq platforms can be effectively applied to neurons in the zebrafish brain.

2.4.1 Graph Clustering Identifies 15 Transcriptionally Distinct Clusters in the Larval Habenula

To identify neuronal types, we used the droplet dataset, because cell type identification was more robust with more cells sequenced at a shallow depth than few cells sequenced deeply (detailed analysis in Figure 2.5) (Shekhar et al., 2016). We used standard computational pipelines to align the raw sequencing data to the zebrafish transcriptome and derive a gene expression matrix of 13,160 genes across 4,233 filtered cells (Methods). To select highly variable genes, we used two approaches: (1) A variable gene selection method implemented in Seurat (Satija et al., 2015), and (2) an alternative approach that ranks genes based on deviation from a null statistical model built on the relationship between variation in transcript counts and mean expression per cell (Methods).

We used principal component analysis (PCA) on a gene expression matrix across 1,436 variable genes and identified 36 statistically significant principal components PCs (n=36, $p < 0.05$). These PCs were used to build a k-nearest neighbor graph of the cells, which was then partitioned into 14 transcriptionally distinct clusters using smart local moving community detection algorithm (Waltman and Eck, 2013) as implemented in the Seurat R package (Satija et al., 2015). Two of these clusters held additional heterogeneity and were further partitioned using iterative clustering (Figure A.1 G). The resulting 16 clusters were visualized in two dimensions using t-distributed stochastic neighborhood embedding (t-SNE) (Figure 2.1 C), and evaluated for differential gene expression to identify cluster-specific markers (Figure 2.1 D) (McDavid et al., 2013; Shekhar et al., 2016; Waltman and Eck, 2013).

Of the 16 clusters identified in our analysis, 15 were habenular neuronal types (Hb01-Hb15) as they robustly expressed *gng8* and other habenular genes (Figure A.2A). We also identified a small sub-group of olfactory placode cells (*Olf*) that are labeled by the *gng8*-GFP line, based on their expression of *epcam*, *calb2a* and *v2rl1* (Figure 2.1E) (Bayramli et al., 2017). This provides an internal control and demonstrates the strength of scRNA-seq to identify contaminant types and remove them from further analysis.

To determine whether clusters corresponded to new or previously described neuronal types, we identified a host of markers for each cluster (Table A1, Figure A.1L). The majority of clusters were defined by newly identified markers (Figure 2.1D). We also determined the relationship between these putative neuronal types using a dendrogram constructed on the variable genes across the dataset (Figure 2.1E). Some clusters recovered in our data comprise <1% of the dataset (Figure 2.1D, side bar), underscoring the advantages of oversampling cells in resolving rare types. Indeed, downsampling analysis on the droplet data suggests that we have reached saturation, but that reducing the number of cells by 20% would lead to recovery of fewer clusters (Figures A.1H-A.1K). These results demonstrate that the small number of neurons in larval habenula can be partitioned into 15 distinct clusters when the cells are sampled at a high frequency.

2.4.2 Most Previously Described Region-specific Genes are Expressed Broadly Across Multiple Neuronal Clusters

To understand how our cluster-specific markers relate to previously described habenular genes, we examined the expression patterns of these spatially localized genes among our clusters (Figures 2.2 and A.2A). We found three major patterns.

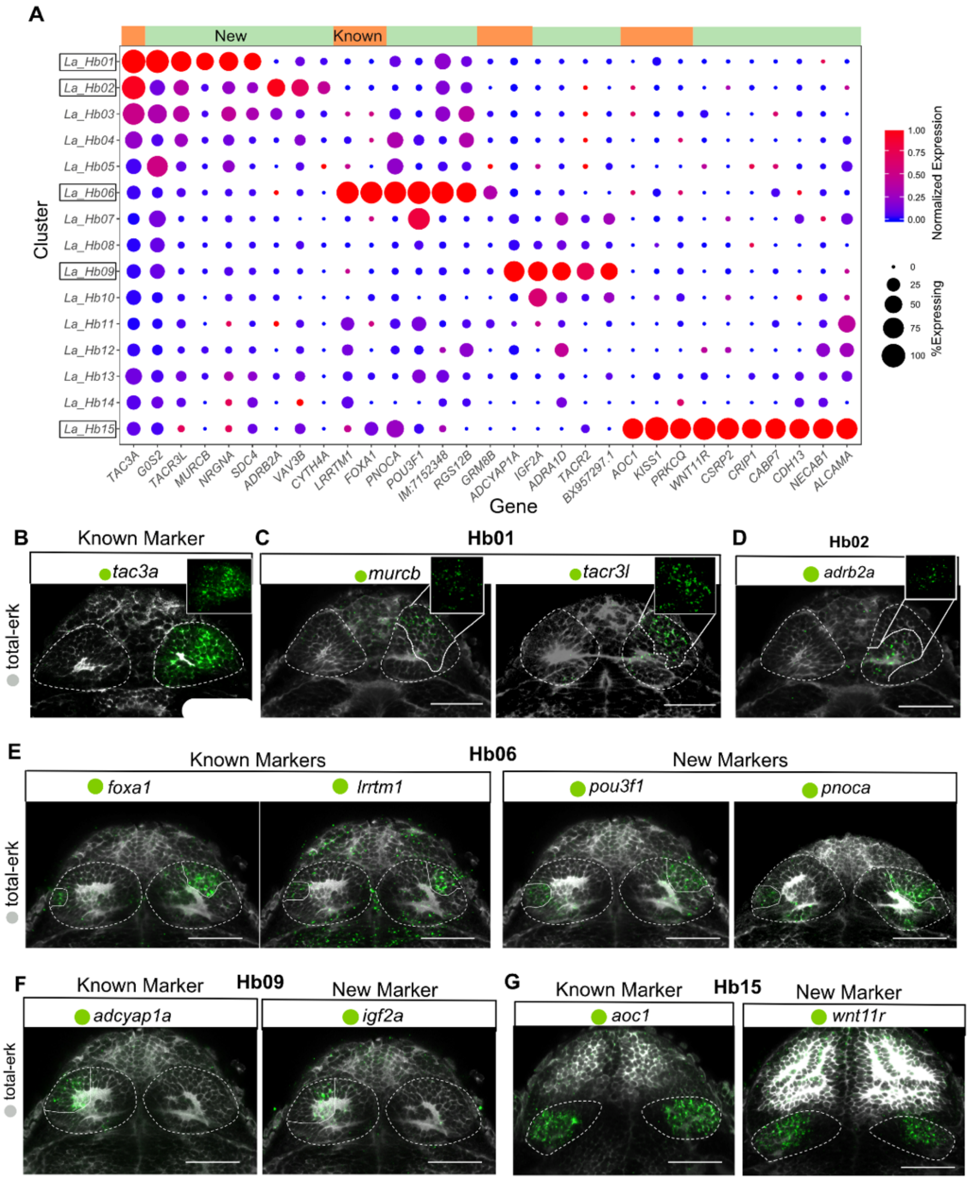


Figure 2.2. Validation and Spatial Distribution of Previously Described Neuronal Types along with Identified Novel Markers.

A) Expression profiles of known and novel habenular marker genes that are specific or enriched in the five clusters displaying previously described gene expression signatures. Green bar on top represents new markers and orange bar represents known markers.

Figure 2.2 (Continued). (B-G) In vivo expression patterns of known and novel marker genes that are enriched in clusters harboring previously characterized habenular genes (Hb01, Hb02, Hb06, Hb09, and Hb15). Each type was characterized by both previously described markers, and new markers found from single cell analysis. RNA-FISH (green) was performed with a total-Erk (pale gray) co-stain for registration. In some cases, a non-linear filter (gamma = 0.3) was applied to the total-Erk (gray) channel to aid visualization of the *in situ* signal (green). B-D) FISH labeling of B) Previously known marker (*tac3a*) and new markers for C) Hb01 (*murcb, tacr3l*) and D) Hb02 (*adrb2a*) found by single cell analysis. Insets show regionalized expression of the gene without total-Erk. *murcb/tacr3l*$^+$(Hb01) and *adrb2a*$^+$(Hb02) domains form subdivisions within the *tac3a*$^+$ domains. E) FISH labeling of new markers *pou3f1* and *pnoca* enriched in the *lrrtm1*$^+$ and *foxa1*$^+$ cluster Hb06. F) FISH labeling of new marker *igf2a enriched in adcyap1a*$^+$ left-only cluster Hb09. G) FISH labeling of new marker *wnt11r* specific to the *aoc1*$^+$ ventral habenular cluster Hb15. Scale bars indicate 50 μm.

First, some known genes show a near perfect overlap with single clusters and were also independently identified as cluster specific markers by our analysis. For example, *adcyap1a*, a known ‘left-only’ marker (Agetsuma et al., 2010) is specific to cluster Hb09. Similarly, known ventral genes such as *aoc1* (Agetsuma et al., 2010; Amo et al., 2010; Hong et al., 2013) *and kiss1* (Kitahashi et al., 2009) are specific to cluster Hb15 (Figure 2.2A). For such neuronal types, we nominated a host of additional markers (*e.g.*, Hb06 (*pnoca, pou3f1*), Hb09 (*igf2a, adra1d, tacr2*), Hb15 (*csrp2, crip1, cabp7, cdh13, wnt11r))* and confirmed their spatially restricted expression patterns (Figures 2.2E-G).

Second, some genes that were known to mark subdomains within the dorsal habenula spanned a few clusters. For instance, *tac3a,* a previously described marker (Biran et al., 2012), is expressed in multiple clusters but at much higher levels in Hb01 and Hb02 (Figure 2.2A). Cluster specific markers for Hb01 (*murcb*) and Hb02 (*adrb2a*) seemed to subdivide the high *tac3a* expressing neurons in the habenula (Figures 2.2A-D).

Third, most genes previously reported to have regional expression patterns within the habenula were expressed across multiple clusters (Amo et al., 2010; Biran et al., 2012; deCarvalho et al., 2014; Kitahashi et al., 2009; Kuan et al., 2007) (Figure A.2A). For example, *gpr151/pou4f1*, which together form classic dorso-medial markers (Broms et al., 2015; Chou et al., 2016) and *nptx2a*, a dorso-lateral marker (Agetsuma et al., 2010), were expressed in majority of clusters, albeit at different levels (Figure A.2A). We detected higher expression of *nptx2a* in the left habenular clusters (Figure A.2A), confirming the previously observed over-representation of lateral identity in the left habenula (Kuan et al., 2007). We validated the overlap of these classic markers with multiple clusters by RNA-FISH (Figure A.2B), demonstrating that individual genes reported to be spatially restricted could span multiple neuronal subtypes.

Among our newly identified markers, we found: 1) markers that are expressed exclusively in the cluster of interest (‘digital’), 2) markers that display 2-3 fold enrichment in the cluster of interest (‘analog’). To compare the specificity of marker genes, we computed their area under the precision recall curve (AUCPR) as a quantitative measure of “cluster-specificity” (See Methods). Digital markers such as *aoc1* displayed AUCPR values greater than 0.8 (Figure A.2C) whereas analog markers such as *tubb5* exhibited lower AUCPR values (0.55 – 0.7). We found that considering pairs of analog markers affords higher specificity in defining a cluster. For instance, *pou3f1* and *pnoca* together localize Hb06 better than each does individually (Figure 2.2E). Therefore, we only used single markers with AUCPR > 0.7 to spatially localize clusters of interest.

Taken together, these results provide novel analog and digital markers for the five clusters that were readily defined by expression of known genes and describe their *in vivo* spatial localization. Strikingly, the majority of previously described habenular genes are broadly expressed over multiple clusters, limiting their utility as markers for transcriptionally distinct neuronal types.

2.4.3 Analysis and Validation of Previously Uncharacterized Habenular Neuronal Types

The remaining 10 of 15 clusters expressed marker genes that to our knowledge have not been previously described in the zebrafish habenula. We hypothesized that these clusters corresponded to potentially novel neuronal subtypes and performed RNA-FISH with cluster-specific markers to validate them and determine their spatial localization (Figure 2.3). We also utilized the gene expression dendrogram (Figure 2.1E) to relate them to ‘known’ neuronal types described earlier. Based on spatial localization, we found three major categories of neuronal subtypes.

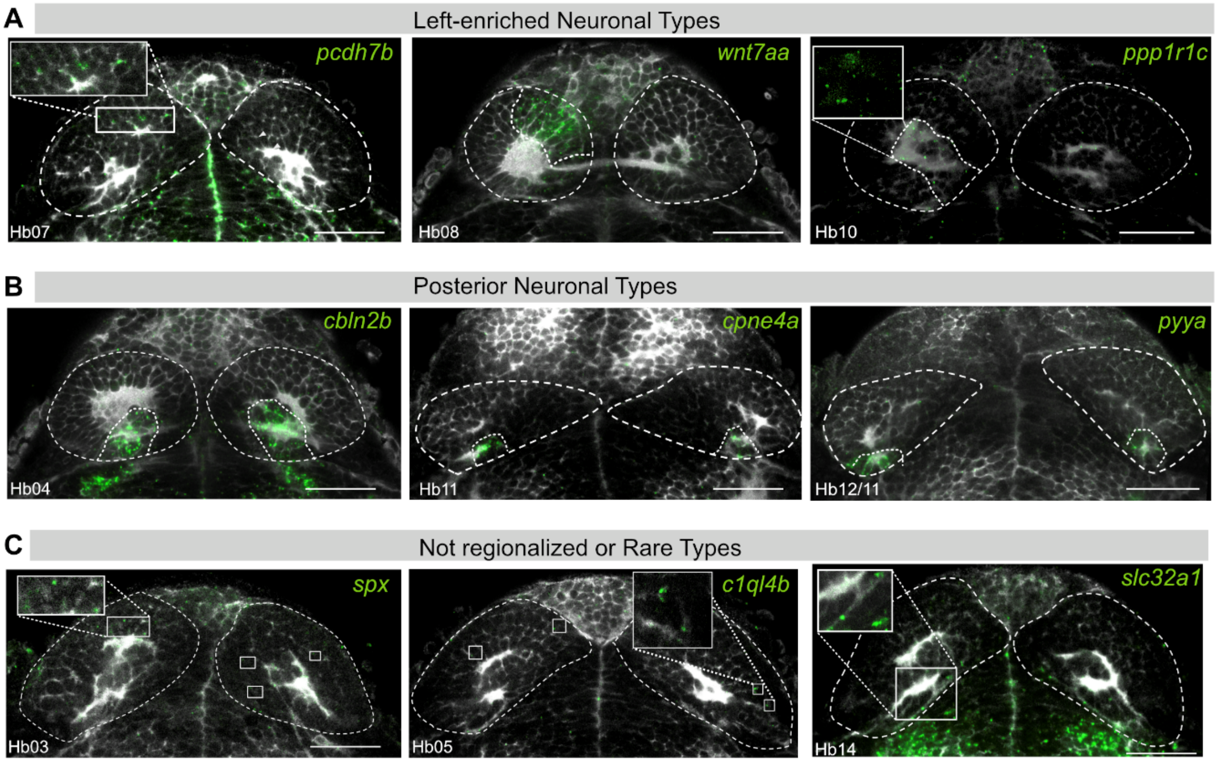


Figure 2.3. Validation and Spatial Distribution of 10 Novel Habenular Neuronal Types.

RNA-FISH (green) was performed for specific markers for novel clusters A) Left-enriched clusters: Hb07 (*pcdh7b*), Hb08 (*wnt7aa*), Hb10 (*ppp1r1c*); B) Posterior habenular clusters: Hb04 (*cbln2b*), Hb11 (*cpne4a*), Hb12/11 (*pyya*), C) Non-regionalized or rare neuronal types: Hb03 (*spx*), Hb05 (*c1ql4b*), Hb14 (*slc32a1*) each overlaid with a total-Erk co-stain (pale gray) for registration. In each case, representative habenular slices with expression are shown.

First, we identified three new 'left-enriched' clusters of neurons – Hb07(*pcdh7b*$^+$) and Hb08(*wnt7aa*$^+$), Hb10(*ppp1r1c*$^+$) – all of which were closely related to one another and to 'left-only' neuronal type Hb09 (*adcyap1a*$^+$) (Figure 2.1E). All of these clusters also expressed another known 'left-only' marker, *nrp1a* (Figure A.2A) (Kuan et al., 2007). Hb08 (*wnt7aa*$^+$) and Hb10 (*ppp1r1c*$^+$) neurons were localized more dorsally in the left habenula than Hb07 *(pcdh7b*$^+$*)* (Figure 2.3A).

Second, we identified three posterior L-R symmetric habenular neuronal types (Figure 2.3B). Dorsally located Hb04 was characterized by the expression of *cbln2b*, a less studied member of the cerebellin genes, some of which are important for synaptic plasticity (Hirai et al., 2005). Ventrally located *pyya*$^+$ neurons were subdivided into Hb11 (*cpne4a*$^+$) and Hb12 (*htr1aa*$^+$). Corresponding to their low proportion by scRNA-seq, RNA-FISH showed that Hb11 and Hb12 are both rare, composed of 4-6 neurons *in vivo*.

Third, we found four rare neuronal types, each comprising less than 5% of the cells in our dataset. Hb03(*spx*$^+$) and Hb05(*c1ql4b*$^+$) form rare populations that seem to be distributed in a non-regionalized manner (Figures 2.3C). Hb13 is a cluster of immature neurons in the medial ventral habenula and lining the ventricular zone, characterized by the expression of *tubb5* and a host of ribosomal proteins (Figure 2.3C) (Ngo et al., 2014). Rarest among these four types were GABAergic neurons in the dorsal habenula (Figures 2.2D and 2.3C) characterized by the expression of *gad1b*, *gad2* and the GABA transporter *slc32a1* and corresponding to 2-3 neurons *in vivo*.

Together, these results provide validation and spatial localization for the 10 novel neuronal types identified by our single-cell analysis and demonstrate a general trend for regionalization of neuronal types within the habenula. Moreover, transcriptional proximity was reflected by spatial proximity in vivo, suggesting that developmental patterning of molecularly related neuronal types occurs in a spatially restricted manner.

2.4.6 Neuronal Types in the Larval Habenula and Their Molecular Signatures are Robustly Reproduced in Full-Length, Deeply Sequenced Libraries

Because zebrafish neurons contain less RNA than other cell types – including of comparable size – analyzed using scRNA-seq (Figures A.1A and A.1B) (Shalek et al., 2013; Shekhar et al., 2016), we asked if we could derive a better classification by sequencing these neurons at a greater depth. To explore this possibility, we prepared 1,152 SMART-seq2 (SS2) libraries using cells sorted from the *gng8*-GFP transgenic line (Picelli et al., 2013), and sequenced them at a median depth of 1 million reads per cell (~25 fold deeper than the droplet data; Figure A.4A and A.4B), resulting in 3,850 genes/cell, (~3 fold more than the droplet data). An independent clustering of 1,040 quality-filtered SS2 cells revealed only 10 clusters (Figure 2.4A), fewer than the droplet data.

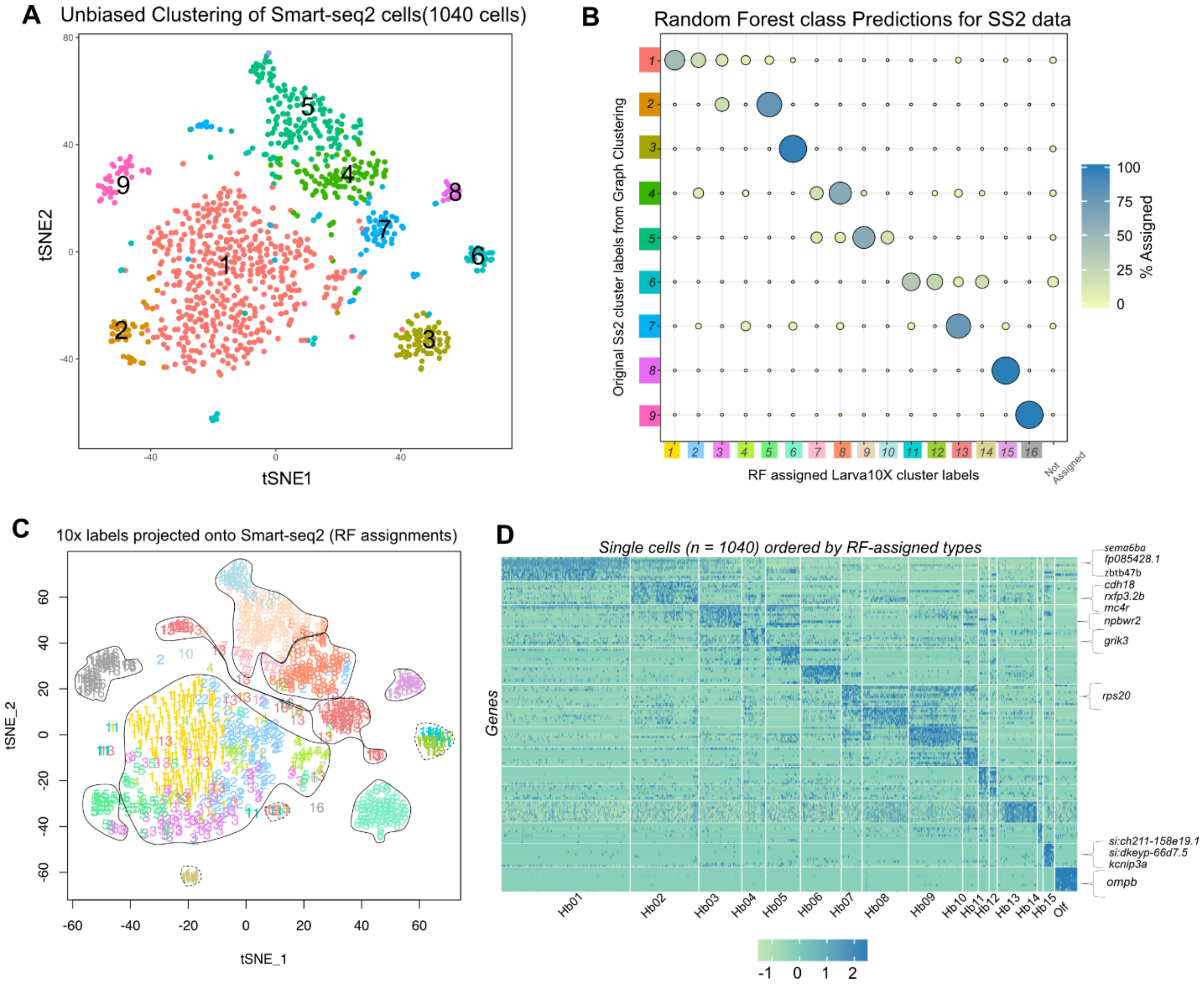


Figure 2.4. Correspondence of Larval Habenular Neuronal Types and their Molecular Identities Between the Droplet and SMART-seq2 Datasets.

(A) t-SNE visualization of single cell clusters obtained by clustering of the SMART-seq2 (SS2) data. (

B) Dot plot (confusion matrix) showing the proportion of cells in each SS2 cluster (rows) that were classified to droplet clusters (columns) using a multiclass random forest classifier (RF). A cell was assigned to a droplet cluster label if > 15 % of the decision trees in the RF classifier contributed to the majority vote (given that there are 16 classes, 6.25% vote would constitute a majority). * represents SS2 clusters in which greater than 70% of the cells of the cluster maps to single droplet clusters.

(C) Same as A, but where each cell is annotated according to its RF assigned droplet cluster label. Rough demarcations of the SS2 clusters as in A are sketched.

(D) Top 10 differentially expressed genes in each habenular type computed using a post hoc test on the SS2 data based on the RF-assigned cluster label as in C. Highlighted on the right are anecdotal examples of genes that were not detected among the top 15 differentially expressed genes in the corresponding droplet clusters.

We hypothesized that the lower number of cells in SS2 dataset led to merging of closely related clusters. To evaluate the correspondence between droplet and SS2 clusters, we trained a multiclass random forest classifier (RF) on the cluster labels of the droplet dataset (Figure A.4I) and used it to map all the SS2 cells onto droplet-based labels (Breiman, 2001; Shekhar et al., 2016). We observed that 5 out of 10 SS2 clusters mapped 1:1 with single droplet clusters (Figure 2.4B). Each of the four remaining SS2 clusters mapped to multiple (typically 2-3) droplet clusters (Figure 2.4B). By labeling each cell on the tSNE plot with RF assigned cluster labels, we observed additional sub-structure in the merged SS2 clusters that was masked in unsupervised clustering (Figure 2.4C). This co-clustering occurs in cases where clusters are closely related (Figure 2.1E) but not highly represented in the dataset (eg: Hb05-Hb03, Hb02-Hb01-Hb04, Hb07-Hb08). These results are also consistent with the recovery of fewer and less pure clusters in the downsampled droplet dataset (Figure A.1H -A.1K).

Next, we asked if the higher number of genes/cell identified in SS2 data enabled the identification of novel cluster-specific markers. To this end, we examined genes that were robustly detected in the SS2 data but not in droplet data (Figures A.4J- A.4K and Figure A.4L, red), but found that they were expressed across multiple clusters and were uninformative for cell type classification (Figure A.4M, Methods). Using differential gene expression analysis on the RF-assigned cell labels, we identified a small number of novel cluster-specific markers that were not identified in droplet dataset (Figure 2.4D, highlighted along heatmap).

Taken together, these results demonstrate a remarkable consistency of types and markers discovered by two different scRNA-seq platforms. Also, consistent with earlier work (Shekhar et al., 2016), these results show that cell type identification is best served by distributing a given number of reads over a large number of cells.

Neurons in the Adult Habenula Retain the Molecular Identity of Larval Types

The habenula undergoes significant growth, morphogenesis and functional maturation from developing larvae to mature adults (Figure A.5I, left). For example, several behaviors mediated by habenular neuronal subpopulations such as aggression are only displayed by adult fish, and the number of neurons increases dramatically between larva and adult (Agetsuma et al., 2010; Chou et al., 2016). Furthermore, the larval dataset was generated with a transgenic line with FACS and may have missed rare populations not labeled by the *gng8*-GFP. To assess the conservation and retention of neuronal types from larva to adult fish (1-year old) and to capture cells that were not labeled by the transgenic line, we dissected whole adult habenulae and performed droplet scRNA-seq (STAR Methods). Post quality filtering, we obtained 7,782 single cell profiles at a median depth of ~96,000 reads and 709 genes per cell (Figure A.5A, A.5B). Using the same clustering approach, we detected 17 clusters and enriched markers (Figures 2.5A and 2.5B) and labeled them post-hoc by comparison to the larval clusters (below).

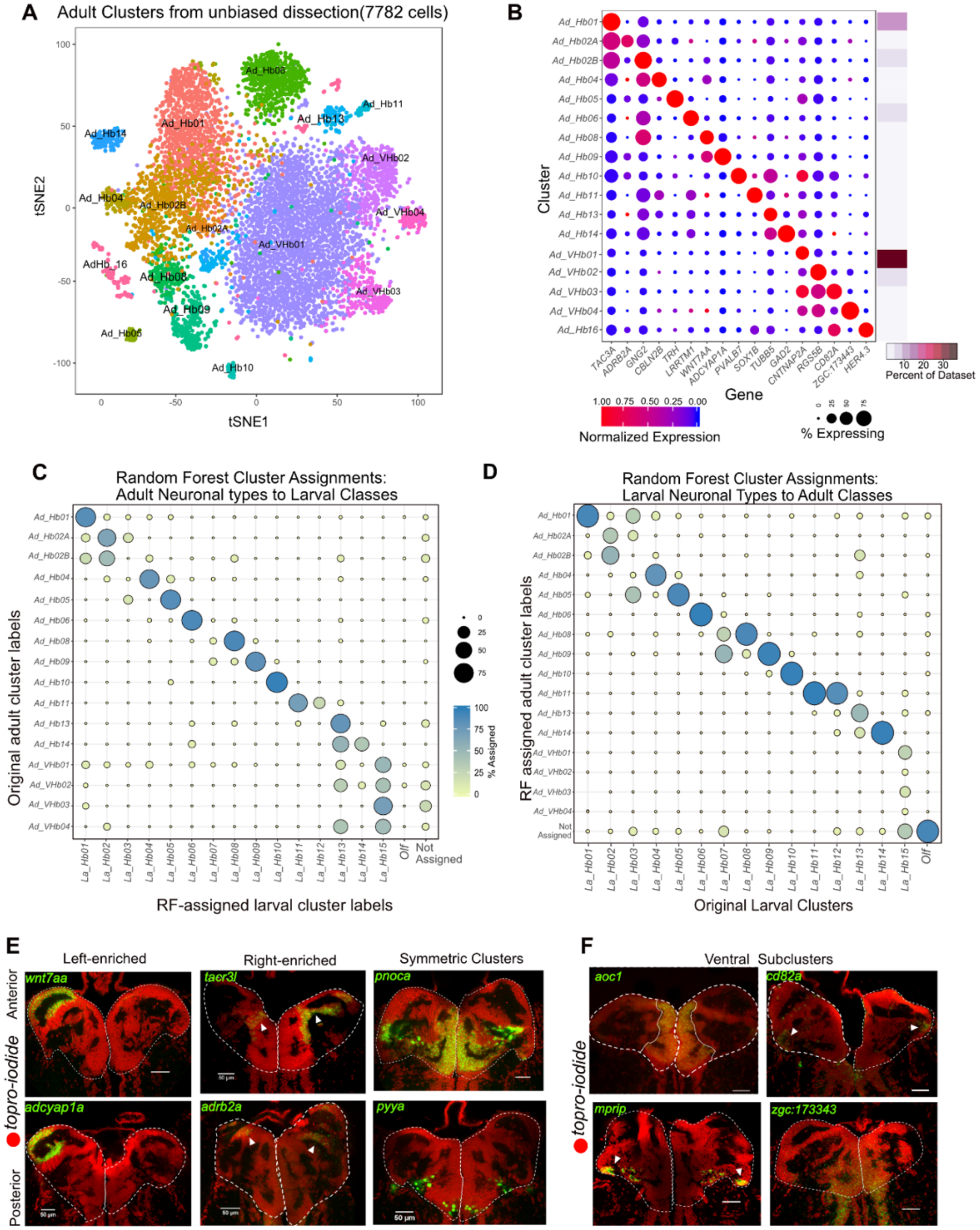


Figure 2.5. Comparative Analysis of Habenular Neuronal Types between Larval and Adult Stages.

(A) t-SNE visualization of adult single cell clusters obtained by clustering of the adult dataset. Clusters have been labeled *post hoc* after comparison to the larval dataset (See Figures 5C and 5D).

Figure 2.5 (Continued). (B) Gene Expression profiles (columns) of select cluster-specific markers identified through differential expression analysis (DEA) across all adult clusters. Bar on the right displays percent of total dataset represented in every adult cluster, showing the abundance of each cell type found by clustering analysis.

(C) Dot plot (confusion matrix) showing the proportion of *gng8*$^+$ cells in the adult dataset (rows) that were classified to larval cluster labels (columns). Each adult habenular type was assigned to a larval cluster label if >15% of the trees in the RF model contributed to the majority vote. Proportion of cells in each row should add to a 100%.

(D) Dot plot (confusion matrix) showing the proportion of larval cells (rows) that were classified to cluster labels of the *gng8*$^+$ cells in the adult dataset (columns). Each adult habenular type was assigned to a larval cluster label if >15% of the trees in the RF model contributed to the majority vote. Proportion of cells in each column should add to a 100%. This training on the adult dataset was performed to validate the robustness of the RF analysis.

(E) FISH validation and localization of select dorsal habenular cluster markers.

(F) FISH validation of the genes that are expressed in all ventral clusters (*aoc1*) and across three other ventral sub-clusters (*cd82a, mprip* and *zgc:173443*).

To systematically compare the clusters between larva and adult, we used a random forest model trained on either dataset to map gene expression signatures between the two datasets (Figure 2.5C (trained on larva) and Figure 2.5D (trained on adult)). Since the larval dataset was generated by cellular sorting based on the *gng8*-GFP line, we restricted our analysis to the high *gng8*$^+$ cells in the adult dataset. Using the RF model, we then classified each high *gng8*$^+$ adult cell (Figure A.4I) into one of the larval cluster labels (Figure 2.5C). Surprisingly, we found that 9 out of the 16 adult habenular clusters mapped 1:1 to single larval habenular clusters (Figure 2.5C). The left-right spatial organization of some clusters was also roughly preserved (Figure 2.5E).

We also found that some larval clusters map to multiple adult clusters. For instance, La_Hb02 (La = Larval) split into Ad_Hb02A and Ad_Hb02B (Ad = Adult). Upon interrogating the differences between Ad_Hb02A and Ad_Hb02B further, we found that certain genes such as *nebl* are selectively expressed in Ad_Hb02B (Figure A.5G) whereas others such as *rac2* are in Ad_Hb02A. Furthermore, larval ventral cluster, La_Hb15 mapped to multiple adult types, all of which are ventral (Ad_VHb01-04) as found by RNA-FISH for differentially expressed markers *cntnap2a*, *mprip*, *cd82a*, and *zgc:173443* (Figure 2.5F). The spatial localization of these ventral clusters was found to be consistent with previously described (Amo et al., 2010) morphogenetic changes that occur in the habenula between larval and adult stages (Figure A.5I).

This multi-mapping among ventral clusters was likely caused by incomplete labeling of ventral habenula by the *gng8-GFP* line, which was used to capture cells in larval dataset. In particular, *in situ* hybridization revealed that two of the adult ventral type markers were expressed in a spatially restricted pattern in the larval ventral habenula (Figure A.5D). We also found a small population of neuronal progenitors (*her4*$^+$, *fabp7a*$^+$, *mdka*$^+$) in the medial ventral habenula (Figures 2.5B and A.5E), in a similar location as the immature *tubb5*$^+$ neurons in the adult dataset.

Taken together, these results demonstrate that despite significant growth and functional maturation between larvae and adults, a substantial proportion of neuronal types in the habenula remain largely constant.

2.4.7 Neuropeptidergic Signaling, Neurotransmission and Neuroexcitability Among Habenular Neuronal Types is Highly Diverse

Previous studies have used ISH or immunostaining to assess the expression of various neuropeptidergic and neurotransmitter genes in the habenula (deCarvalho et al., 2014; Hong et al., 2013) but a comprehensive expression profile of these genes is not available. Using the scRNA-seq dataset, we assessed the expression profiles of these genes among the different neuronal types in the larval (Figure 2.6) and adult (Figure 2.7) habenula.

We found that a large number of neuropeptides are specific to a small number of neuronal types (Figures 2.6A and 2.7A). For example, *spx, trh, adcyap1a, tac1, agrp, kiss1, npy, galn, penka* were all expressed in only one habenular type each, whereas *pyya, pyyb, pdyn, pnoca, sst1.1, tac3a, cckb* were enriched in 2-4 neuronal types. In addition, co-expression of multiple neuropeptides was a common feature among habenular neuronal types. For instance, *spx*, which acts as a satiety factor and *pyyb,* which has been implicated to be an anorexigenic factor, were co-expressed in the homologous clusters in both larvae and adults respectively (Wong et al., 2013). Conversely, we examined the expression profile of all detectable neuropeptide receptors. In some cases, the same neuronal type produces a neuropeptidergic signal and expresses its cognate receptor, suggesting autocrine signaling. For instance, *kiss1* and its receptors *kiss1rb* and *kiss1ra* are co-expressed in ventral habenular clusters (Figures 2.6A and 2.7A).

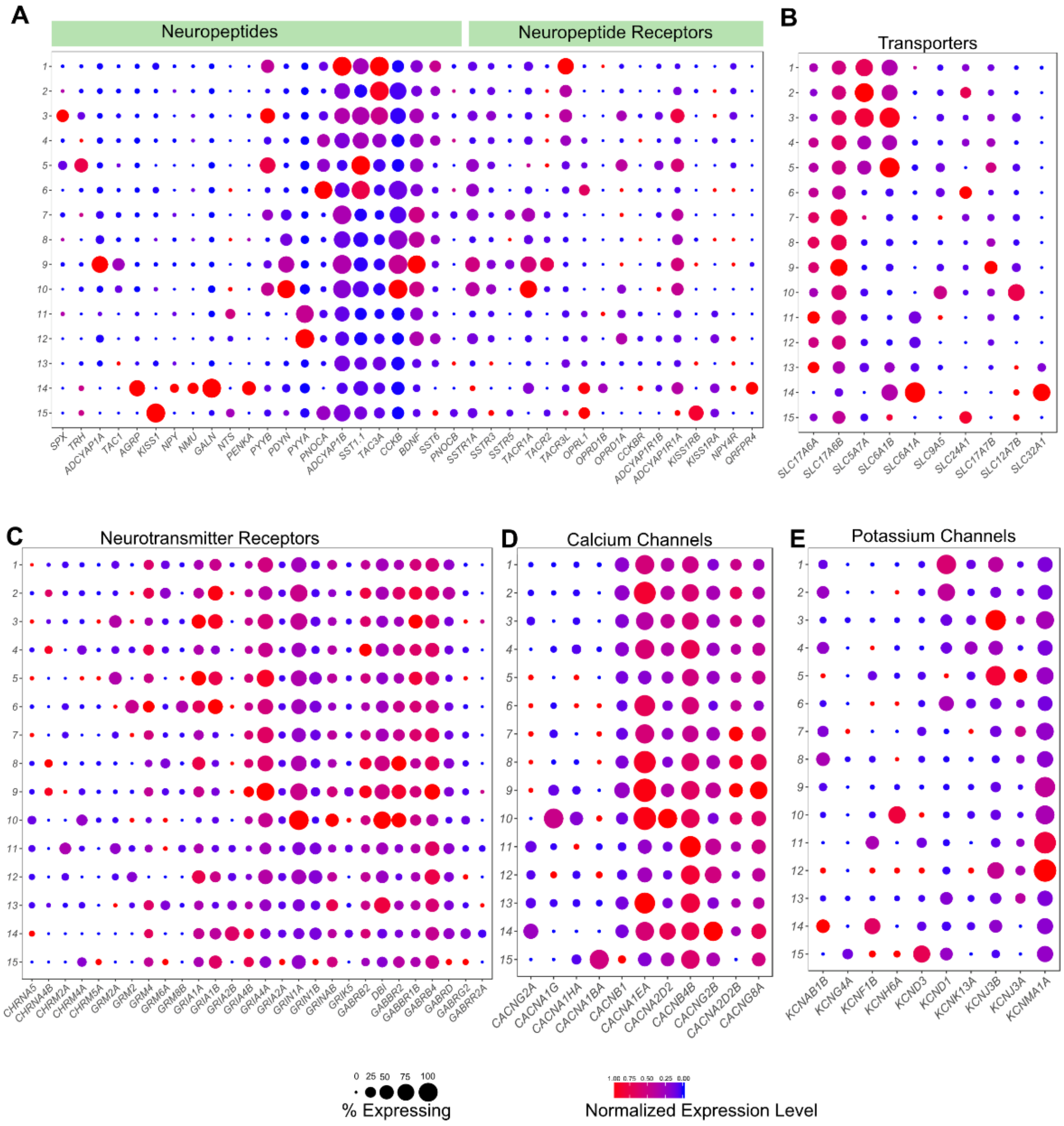


Figure 2.6: Divergent Expression Patterns of Functionally Relevant Genes among the Larval Habenular Neuronal Types.

(A-F) Gene expression profiles of select functionally relevant genes among larval habenular types visualized in the form of a dot plot. Representation as in Figure 1D. A) Neuropeptides and Neuropeptide Receptors B) Transporters C) Neurotransmitter Receptors D) Calcium Channels E) Potassium Channels. Only genes expressed in >20% of cells in at least each habenular type are shown.

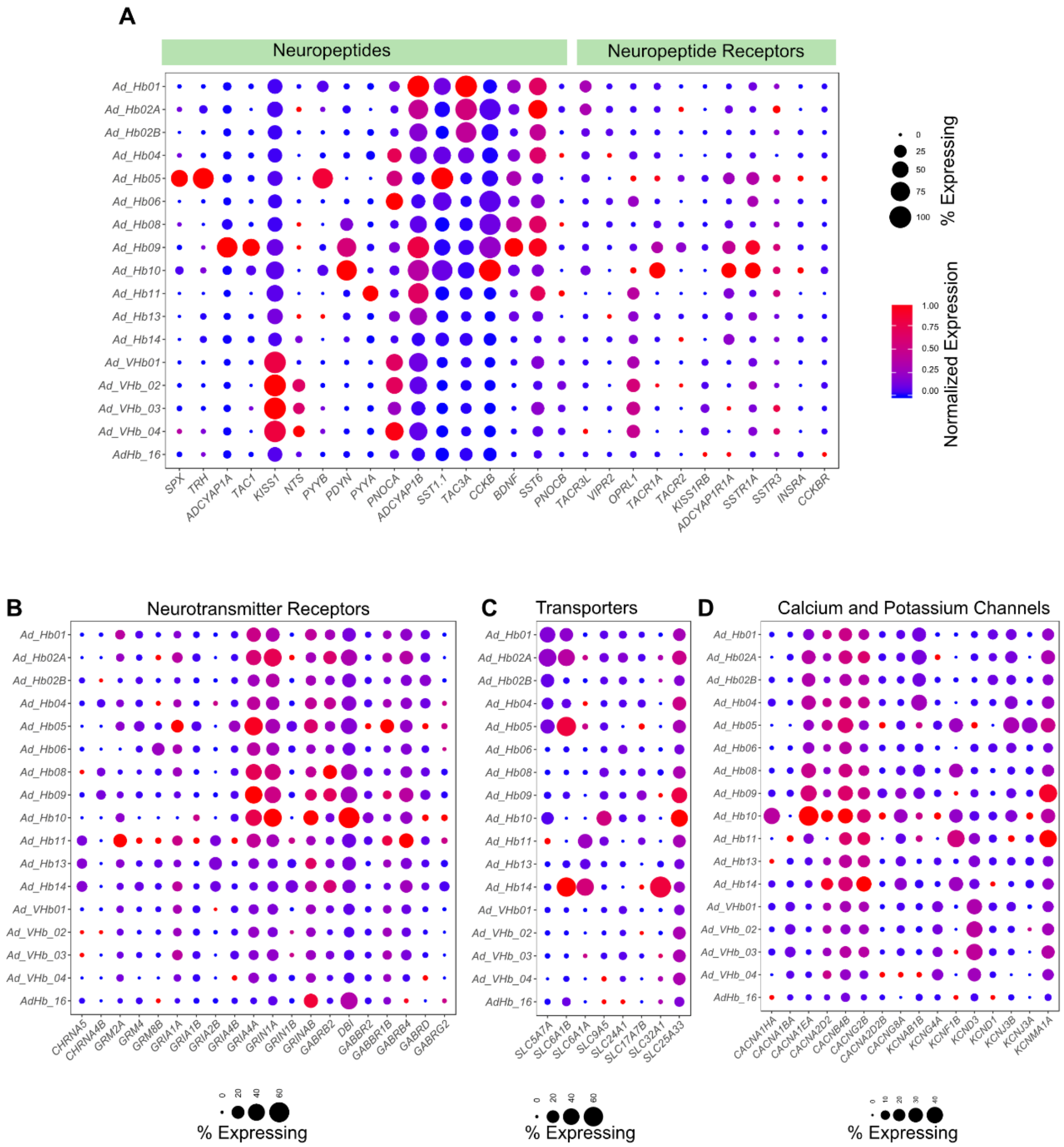


Figure 2.7. Divergent Expression Patterns of Functionally Relevant Genes among the Adult Habenular Neuronal Types.

(A-F) Gene expression profiles of select functionally relevant genes among adult habenular types visualized in the form of a dot plot. Representation as in Figure 2.6. A) Neuropeptides and Neuropeptide Receptors B) Neurotransmitter Receptors C) Transporters D) Calcium and Potassium Channels.

Next, we analyzed the expression of genes essential to signal transduction in neurons. As previously observed (Appelbaum et al., 2009), a large proportion of habenular neurons are glutamatergic, as reflected in the broad expression of glutamate transporters *slc17a6a* and *slc17a6b* (Figure 2.6B). However, the habenula also contains a small population of GABAergic neurons (Figure 2.6B). Hb01 through Hb05 are also cholinergic and express the choline transporter *slc5a7a* (Figure 2.6B). Cholinergic transmission was largely absent in left-enriched neuronal types consistent with previous studies (Hong et al., 2013). Similarly, some neurotransmitter receptor subunits such as *grm8b, grm2, gria2b* are enriched in one or a few clusters (Figure 2.6C). Calcium channels such as *cacng2a, cacna1g, cacna1ba* are also expressed in a type specific manner, as are potassium channel subunits such as *kcnab1b, kcnf1b, kcnh6a, kcnd3* (Figure 2.6D and E). We found a similar distribution of these functionally relevant genes in the adult habenula (Figure 2.7A-F). Collectively, our data shows that despite being a small region, the habenula shows a remarkable diversity in expression of genes that directly influence the functional and electrophysiological properties of individual neurons.

2.5 DISCUSSION

We used two scRNA-seq platforms, integrative computational analysis, and brain registration to build and validate a comprehensive atlas of neuronal types in the larval and adult zebrafish habenula. Our study provides six main advances. First, we devised a robust protocol for the dissociation and capture of single neurons from the zebrafish brain. Second, we found that comprehensive identification of neuronal types by scRNA-seq can be achieved by high cell-sampling coverage of a small brain region. Third, we discovered thirteen new neuronal types and identified fine-grained spatial subdivisions in the habenula. Fourth, we discovered dozens of new marker genes that define habenular neuronal types. Fifth, we found that diverse neuronal types are largely retained from larva to adult. Sixth, we showed that the reference atlas enables comparison of molecularly defined neuronal types with those defined by neural activity. Taken together, our study creates a resource for future studies on habenular development and function and provides a technological framework for the characterization of other brain regions.

Single-Cell Analysis of the Zebrafish Brain

Although scRNA-seq is now well-established, applying it to zebrafish neurons presents certain technical challenges. Our work establishes critical conditions for future scRNA-seq in the zebrafish brain. We provide a robust dissociation and sorting based cell capture protocol for zebrafish neurons (Methods), which were found to have comparatively less RNA than other cell types. Consistent with previous studies, we also found that transcriptional signatures required to classify cell types can be identified by low coverage RNA-seq (Shekhar et al., 2016). Therefore, most of the additional genes detected in the deeper SMART-seq2 dataset were uninformative for cell type classification (Figure A.4M).

Comprehensive Cell Type Identification and Spatial Mapping in the Habenula Facilitates Functional Studies

Based on scRNA-seq on larva and adults, we found that the habenula is composed of at least 18 distinct neuronal subsets. While larval ventral heterogeneity needs to be explored genome wide using RNA-seq approaches, our two-time point scRNA-seq and *in situ* analysis has allowed us to describe the majority of the neuronal types in the habenula. For each of these neuronal types, we found dozens of novel molecular markers. Whole brain *in situ* hybridization showed that some of these markers are exclusively expressed within habenular neuronal subsets, making them good candidates for generating reporter lines (Figure A.3C).

We also used RNA-FISH and image registration to spatially localize neuronal types, a key step in linking molecular profiles to physiological and behavioral features. We found that 15 neuronal subsets are highly regionalized to habenular sub-regions (Figures 2.3L, 2.5F). This arrangement is similar to regionalization observed in the hippocampus (Habib et al., 2016) but in contrast to the organization in the retina where different cell types are intermixed within the same spatial location (Masland, 2001) .

Furthermore, we found that previously demarcated anatomical sub-regions, dHbM (*pou4f1*$^+$, *gpr151*$^+$), dHbL (*nptx2a*$^+$) and vHb (*aoc1*$^+$) harbor multiple transcriptionally distinct neuronal subsets. Previous studies have shown that these distinct sub-regions send efferent projections to distinct anatomically separated sub-regions of the downstream interpeduncular nucleus (IPN) and raphe nucleus (Beretta et al., 2012). Our study raises the possibility of a finer topography of efferent projections of transcriptionally distinct neuronal subsets into finer sub-regions of downstream targets.

We also found a large diversity of neuropeptidergic genes in the habenula, which may shed light into its unexplored functional roles. For instance, a number of peptides that are involved in food intake regulation such as *cckb, pyya, pyyb,* and *spx* are shown to be robustly expressed in certain sub-clusters of habenular cells, indicating that the corresponding neuronal types may be important for regulating food intake. Furthermore, a subset of the *pyya* neurons are positive for *htr1aa*, a serotonin receptor whose expression is known to be affected by anorexigenic drugs (Shimada et al., 2012).

Therefore, our comprehensive list of neuronal types along with dozens of marker genes and their spatial map will be a valuable resource for the study of habenular development and function in normal and pathophysiological conditions.

Neurons in the Adult Habenula Retain the Molecular Identity of Larval Types

Cell type identification should distinguish between stable cell types and transient cell states as transcriptional cascades in individual neurons may change in response to a variety of different stimuli such as neural activity, neuropeptides, hormones or developmental signals. Our two time-point study design enabled us to observe the remarkable congruence of habenular subtypes between the developing 10-day old larval and the adult brain, suggesting that these signatures represent stable molecular identities. We also found that the localization of neuronal types in the adult habenula is consistent with the complex morphogenetic changes in the habenula between larva and adult wherein the ventral cells migrate inward from lateral to medial positions (Figure A.5H, A.5I) (Amo et al., 2010).

A number of habenula-dependent behaviors such as fear responses and aggression arise later in development in juvenile or adult zebrafish (Agetsuma et al., 2010; Chou et al., 2016; Dreosti et al., 2015). In the absence of major changes in molecular cell type diversity, what circuit mechanisms account for alterations in behavioral capacities during development? An attractive hypothesis is that habenular input-output relationships change between the two time-points. For example, habenular outputs may diverge into more downstream regions to mediate a diverse set of behaviors in adults. A detailed investigation of the development of habenular inputs and outputs will be required to address this question.

Pipeline for Comprehensive Identification of Cell Types in Other Brain Regions

The comprehensive classification of neural types is limited by the large number and diversity of cells in vertebrate brains. For example, previous scRNA-seq studies have sampled a small percent of cells to infer heterogeneity in large regions (Campbell et al., 2017; Chen et al., 2017; Gokce et al., 2016; Zeisel et al., 2015). Our study shows that such limitations can be overcome by scRNA-seq from predefined small sub-regions at high cellular coverage. The larval zebrafish brain contains fewer than 200,000 neurons, putting its comprehensive molecular classification within reach of current technologies and creating a blueprint for cell type diversity in the vertebrate brain.

2.6 MATERIALS AND METHODS

Experimental Model and Subject Details

Larvae and adult fish were maintained on 14 hours: 10 hours light: dark cycle at 28ºC. All protocols and procedures involving zebrafish were approved by the Harvard University/Faculty of Arts & Sciences Standing Committee on the Use of Animals in Research and Teaching (IACUC; Protocol #25-08). Wildtype zebrafish from the TLAB strain were used. 10 days post fertilization (dpf) larval and ~1-year old adult male and female zebrafish were used across different experiments. Animals were anesthetized in 0.2% tricaine and rapidly euthanized by immersion in ice water for 5 minutes before dissection. RNA-seq experiments were performed with the transgenic lines: Tg(gng8:*nfsB*-CAAX-GFP) and Tg(gng8: GAL4 X UAS: mCherry) in larval and adult stages.

Cell Isolation and RNA-Seq

Isolation of cells for SMART-seq2

10 dpf larval heads were dissected in Neurobasal (ThermoFisher Scientific 21103049) supplemented with 1x B-27 (ThermoFisher Scientific 17504044), and promptly dissociated using the Papain Dissociation Kit (LK003150) with following modifications. Larval heads were incubated in 20 units/mL papain for 25 minutes at 37ºC. The cells were dissociated by gentle trituration 20 times and spun at 300xg for 5 minutes. The cells were resuspended in 1.1 mg/mL papain inhibitor in Earle's Balanced Salt Solution (EBSS). The resulting cell suspension was passed through 20 µm cell strainer and placed on ice. Two viability indicators (calcein blue, a live stain (ThermoFisher Scientific C1429) and ethidium homodimer, a dead cell stain (ThermoFisher Scientific E1169)) were added at a concentration of 0.01mg/mL. If the viability of cells was greater than 85% as measured by calcein blue staining, the cells were immediately sorted directly into a 96-well plate with 5µL of lysis buffer comprised of Buffer TCL (QIAGEN 1031576) plus 1% 2-mercaptoethanol (Sigma 63689). All samples were immediately frozen in dry ice and stored at -80ºC until further processing.

SMART-seq2 library preparation

For preparation of SMART-seq2 libraries, the plates containing single-cell lysates were thawed on ice and purified with 2.2X RNAClean SPRI beads (Beckman Coulter Genomics) without elution. The beads were then air-dried and processed promptly for cDNA synthesis. SMART-seq2 was performed using the published protocol (Picelli et al., 2013) with minor modifications (Shekhar et al., 2016). We performed 22 cycles of PCR for cDNA amplification. The resulting cDNA was eluted in 10μL of TE buffer. We used 1.25 ng of cDNA from each cell and one fourth of the standard Illumina Nextera XT reaction volume in tagmentation, and final PCR amplification steps. We pooled 384 single-cell libraries from *gng8*-GFP line in each batch and sequenced 50x25 paired end reads using a single kit on the Illumina NextSeq500 instrument.

Cell isolation and library preparation for Droplet scRNA-seq

10dpf larval heads from 25 *gng8*-GFP fish were dissected, and cells were dissociated as described previously. After reconstitution of cells with papain inhibitor, the cells were spun again at 300xg for 5 minutes and washed with 1X Phosphate Buffered Saline (PBS). Viability indicators were added and roughly 6000-8000 cells were immediately sorted into 20μL of PBS to a final concentration of cells 300 cells/μL. The sorting was performed with a MoFlo Astrios (Beckman Coulter) with highly reduced sort pressure at 20psi. This was found to be a critical step as higher sorting pressure led to high cell death post sorting. The resulting single cell suspension was promptly loaded on the 10X Chromium system (Zheng et al., 2016). The sorted cells were not kept on ice for longer than 10 minutes because the viability post cell sorting, as measured by trypan blue staining, was found to drop over time.

For experiments with adult animals, six adult habenulae were directly dissected out of the brain based on the expression of *gng8*-GFP marker. The resulting tissue was then dissociated in 1mL of 20 units/mL papain for 15 minutes at 37ºC. The habenular cells were dissociated by trituration, spun at 300xg for 5 minutes and resuspended in 1mL 1.1mg/mL papain inhibitor solution. The resulting cells were then washed in 1x PBS + 200mg/mL Bovine Serum Albumin (BSA) (NEB, B9000S) once and filtered through a 20 μm cell strainer. The cells were then resuspended in 50µL PBS + 200µg/mL BSA and counted on a hemocytometer. Viability was assessed by using a trypan blue staining and cells were loaded onto the 10X Chromium system at a concentration of ~300 cells/µL after ensuring that the cell viability in the suspension was greater than 80%. 10X libraries were prepared as per the manufacturer's instructions (Zheng et al., 2016).

Imaging Methods

Probe synthesis for RNA in situ hybridization

Fragments of the following genes were amplified using Phusion Hi-Fidelity polymerase (New England Biolabs, M0530L) with the primers listed in Table A.3. The Polymerase Chain Reaction (PCR)-amplified fragments were then cloned into pSC-A plasmid using Strataclone PCR Cloning Kit (Agilent, 240205), and used to transform the Strataclone competent cells. The transformed cells were plated overnight on Luria-Bertani (LB) agar plates. Colonies were selected by colony-PCR, cultured, mini-prepped and sent for sequencing. The resulting plasmids were then restricted with the appropriate restriction enzyme (Table A.3), and purified using PCR-clean up kit (Omega Cycle Pure Kit). The linearized vector was then used as a template to synthesize digoxigenin- or fluorescein-labeled RNA probes using the RNA labeling kit (Roche). The transcription reactions were purified using Total RNA clean up kit (Omega, R6834), and the resulting RNA was quantified using Nanodrop and assessed on an agarose gel. The final product was then normalized to a concentration 50ng/μL in HM+ buffer (50% formamide, 5X Saline Sodium Citrate (SSC) buffer, 5 mgmL^{-1} torula RNA, 50 μgmL^{-1} heparin, 0.1% Tween 20) and stored at -20ºC until further use.

Fluorescent in situ hybridization

Fluorescent RNA in situ hybridizations were performed as previously described (Ronneberger et al., 2012). Zebrafish larvae were grown until 10dpf and were fixed in 4% formaldehyde (Sigma-Aldrich) in PBS at 4ºC overnight. Post fixation, the larvae were rinsed three times in PBST (PBS with 0.1% Tween 20), and subsequently dehydrated in increasing concentrations of methanol (10 minutes each 25% methanol: 75%PBST, 50% methanol: 50% PBST, 75% methanol: 25% PBST and two times 100% methanol). Dehydrated larvae were then stored at -20ºC at least overnight. Larvae were rehydrated with decreasing concentrations of methanol (10 minutes each, 75% methanol: 25% PBST, 50% methanol: 50% PBST, 25% methanol: 75% PBST, four times for 10 minutes PBST). They were digested in Proteinase K (10 μg/mL) for 1 hour and immediately fixed in 4% formaldehyde to stop digestion (20 minutes). Larvae were then pre-hybridized in Hybridization Mix (HM)+ buffer (50% formamide, 5× SSC buffer, 5 mg/mL torula RNA, 50 μg/mL heparin, 0.1% Tween 20) at 65ºC for 2 hours. Hybridization reactions with the RNA probes were carried out in HM+ (with digoxigenin-labeled antisense probes for single *in situ hybridizations* and digoxigenin- or fluorescein-labeled antisense probes for *double in situ hybridizations*) overnight at 65ºC. Probes were normalized to a concentration of 3.33ng/µL and denatured at 70ºC for 10 minutes before hybridization. Sense probe controls were performed alongside the antisense probes.

The next day, larvae were washed several times at 65ºC (20 minutes in hybridization mix, 20 minutes in 75% formamide: 25% 2xSSCT (2XSSC with 0.1% Tween20), 20 minutes in 50% formamide: 50% SSCT, 20 minutes in 25% formamide: 75% SSCT, twice for 20 minutes in 2x SSCT, thrice for 30 min in 0.2x SSCT. The larvae were then washed twice in TNT (100mM Tris-HCl, pH 7.5, 150mM NaCl, 0.5% Tween 20) at room temperature. The larvae were subsequently blocked in 1% Blocking Reagent in TNT (TNTB) for at least 1 hour and incubated with a peroxidase-conjugated anti-digoxigenin-POD antibody (1:400 dilution in TNTB) at 4°C overnight with gentle agitation (Anti-Digoxigenin-POD Fab Fragments, Roche 11 207 733 910).

The following morning, the antibody was removed and larvae washed in TNT (8 times for 15 minutes each). After the washes, the larvae were stained per the TSA kit instructions for 1 hour in darkness without agitation (Perkin Elmer TSA Plus Cyanine 3 System, NEL744001KT). The larvae were then washed in TNT three times, 5 minutes each. For single *in situs*, a subsequent immunostaining for anti-total-Erk (Cell Signaling, 9102) was performed to use as an anatomical reference and signal for brain registration across multiple fish.

For double *in situs*, after the first staining reaction, the first antibody was removed by treating the larvae in 1% hydrogen peroxide for 20 minutes without agitation. The larvae were then incubated overnight with anti-fluorescein-HRP antibody (Anti-Fluorescein-POD Fab Fragments, Roche 11 426 356 910), diluted in 1:400 in blocking buffer at 4C with gentle agitation.

The following morning, the antiserum was removed and discarded, and excess antibody was removed by rinsing the embryos 8x15 minutes in TNT buffer. They were then subsequently stained by incubating in 100μL of Cy3 tyramide reagent diluted in 1:25 in amplification diluent (Perkin Elmer TSA PLUS Cyanine 3 System, NEL744991KT) for an hour without agitation. The embryos were then washed in 8X15 minutes in PBST and subsequently stained with anti-total-erk antibody for anatomical reference.

Fluorescent RNA in situ hybridization in adult zebrafish

Adult fish brains were dissected in ice-cold PBS and fixed overnight in 4% paraformaldehyde (PFA). Fluorescent RNA in situ hybridization in adult dissected brains was performed with the same protocol as outlined above with the following changes. Brains were digested in Proteinase K (20ug/mL) for 35 minutes. Following probe hybridization, antibody incubation and tyramide signal amplification, the brains were mounted in 3% low-melt (LM) agarose and sliced into 50 micron sections using the vibratome. The resulting slices were stained with TOPRO3 (1:5000) or Sytox green

(1:30,000) for nuclear staining and imaged using a Zeiss inverted Confocal microscope with a 20X air objective and a 63X oil dipping objective.

Imaging and Image Registration

The larvae were washed three in PBST, mounted in 2% LM agarose and imaged with an upright Confocal Zeiss LSM 880 with a water dipping 20x objective.

Whole Habenula Registration

Image registration across multiple habenulae was performed with CMTK (http://www.nitrc.org/projects/cmtk/) with command string (-T 32 -awr 010203 -X 52 -G 80 -R 3 -A '--accuracy 0.8' -W '--accuracy 1.6'). Template habenula was a 10dpf nacre (mitfaa -/-) larvae that underwent RNA-FISH and immunostaining with anti-total-erk. All RNA-FISH images were subsequently registered to the same reference using cmtk (Jefferis et al., 2007; Randlett et al., 2015; Rohlfing and Maurer, 2003). All registered images were compared to the original *in situ* images to screen out unnatural morphing artifacts. The best registered images were chosen manually and used for generating reference habenula (Movie S1)

Whole Brain Registrations

Whole brain registrations were performed in a similar manner by choosing a reference whole brain image from a larva that underwent RNA-FISH and immunostaining with

anti-total-erk. However, subsequent registrations were performed using Advanced Normalization Tools
ANTs (Marquart et al., 2016) with the following parameters:

```
antsRegistration -d 3 --float 1 -o [fish1_,fish1_Warped.nii.gz] --interpolation
WelchWindowedSinc --use-histogram-matching 0 -r [Ref1.nii,fish1- 01.nii.gz,1] -t
rigid[0.1] -m MI[Ref1.nii,fish1-01.nii.gz,1,32,Regular,0.25] -c [200x200x200x0,1e-8,10] --
shrink-factors 12x8x4x2 --smoothing-sigmas 4x3x2x1vox -t Affine[0.1] -m MI[ref/terk-
ref.nii,fish1-01.nii.gz,1,32,Regular,0.25] -c [200x200x200x0,1e-8,10] --shrink-factors
12x8x4x2 --smoothingsigmas 4x3x2x1vox -t SyN[0.1,6,0] -m CC[ref/terk-ref.nii,fish1-
01.nii.gz,1,2] -c [200x200x200x200x10,1e-7,10] --shrink-factors 12x8x4x2x1 --
smoothingsigmas 4x3x2x1x0vox
```

Due to the high computation time required for these analyses, registrations were parallelized using Slurm-based bash scripts.

Image visualization

A non-linear gamma filter (ImageJ (Math; Gamma = 0.3)) was applied to the total-Erk channel in some images presented in the main text to aid visualization of the FISH signal.

Computational Methods for Data Analysis

Alignment and quantification

For the 10X droplet data, raw sequencing data was converted to matrices of expression counts using the cellranger software provided by 10X genomics[1]. Briefly raw BCL files from the Illumina NextSeq or HiSeq were demultiplexed into paired-end, gzip-compressed FASTQ files for each channel using “cellranger mkfastq”. Both pairs of FASTQ files were then provided as input to “cellranger count” which partitioned the reads into their cell of origin based on the 16bp cell barcode on the left read. Reads were aligned to a zebrafish reference transcriptome (ENSEMBL Zv10, release 82 reference transcriptome), and transcript counts quantified for each annotated gene within every cell. Here, the 10-base pair unique molecular identifier (UMI) on the left read was used to collapse PCR duplicates, and accurately quantify the number of transcript molecules captured for each gene in every cell. Both cellranger mkfastq and cellranger count were run with default command line options. This resulted in an expression matrix (genes x cells) of UMI counts for each sample.

For SS2 data, raw reads were mapped to a zebrafish transcriptome index (Zv10 Ensembl build) using Bowtie 2 (Langmead and Salzberg, 2012), and expression levels of each gene was quantified using RSEM (Li and Dewey, 2011). We also mapped the reads to the Zv10 genome using Tophat2. We only used libraries with genome

---

[1] https://support.10xgenomics.com/single-cell-gene-expression/software/ pipelines/latest /what-is-cell-ranger

alignment rate > 90% and transcriptome alignment rate (exonic) > 30%. RSEM yielded an expression matrix (genes x samples) of inferred gene counts, which was converted to TPX (transcripts per 10^4) values and then log-transformed after the addition of 1, consistent with the normalization of the droplet data.

Filtering expression matrix and correcting for batch effects

Cells were first filtered to remove those that contain less than 500 genes detected and those in which >6% of the transcript counts were derived from mitochondrial-encoded genes (a sign of cellular stress and apoptosis). Genes that were detected in less than 30 cells were also removed. Among the remaining cells, the median number of UMIs per cell was 2,279 and the median number of genes was 1,319 for larval data. The same for adult data was 1,614 UMI/cell and 709 genes/ cell, respectively (Figure A.1C, A.1D, A.5A and A.5B).

We used a linear regression model to correct for batch effects in the gene expression matrix using the RegressOut function in the Seurat R package, and used the residual expression values for further analysis. The residual matrix was then scaled, centered and used for the selection of variable genes, PCA and clustering.

Finding variable genes

To select highly variable genes in the data, we use two methods. For UMI-based droplet data, we derived a null mathematical model based on physical principles to model the relationship between average counts and the coefficient of variation (CV) across all the genes based on a negative binomial distribution (see below). This null model accurately estimated the minimum CV as a function of the mean counts across the full range, such that the actual CV for every gene was larger than the value predicted by our model. We used this model to rank genes based on the “excess CV” (difference between observed and predicted) and identified 706 highly variable genes (Figure A.1F). This model fit droplet data very well (Figure A.1E and A.1F).

However, this model greatly underestimated the CV as a function of mean counts in SMART-seq2 data (Figure A.4D and A.4E), which was better captured by a similar, but more flexible mean-variance model developed earlier for SMART-Seq scRNA-seq data (Brennecke et al., 2013). We speculate that this additional overdispersion in SMART-Seq data is due to amplification biases in the read counts in SMART-seq like protocols, which are not attenuated without UMIs. A similar analysis of droplet data with the raw read counts instead of UMIs supports the hypothesis that amplification biases are responsible for the overdispersion (Figure A.4F). (see details in Section “*Mean-CV model for transcript counts in UMI based data”* below).

We also used Seurat’s data driven variable gene selection (FindVarGenes) method to identify highly variable genes in the SMART-seq2 data (Satija et al., 2015). Briefly, the mean expression and dispersion (variance/mean) for each gene is computed across all single cells. The genes were then placed in 20 bins based on their mean expression. Within each bin, the dispersion of all genes was z-normalized to identify the genes that were highly variable within a similar average expression value. We identified 1,258 genes by this method. For our final analysis of the droplet dataset, we used a union of the variable genes selected with the two methods, which resulted in a total of 1436 variable genes. However, for SMART-seq2 data, which was generated without UMIs, we only used Seurat's dispersion method for variable gene selection.

Mean-CV model for transcript counts in UMI based data

The coefficient of variation (CV), defined as the ratio between the standard deviation (σ) of a variable and its mean (μ), is a natural measure of a gene's extent of variation. However, ranking genes based on decreasing expression CV leads to selection of genes with low mean expression, particularly in count-based data. Hence, we sought to perform variable gene selection in droplet dataset using mean-CV relationship. The simplest null model is that the transcript counts follows the Poisson distribution,

$$X_g \sim Poisson(\mu_g) \qquad (1)$$

where X_g is the UMI counts for gene *g* in a cell, and μ_g is the sampling rate equal to the average count of gene *g* across all the cells. Since the variance of the Poisson distribution is equal to mean, this predicts a relationship: $CV_g = 1/\sqrt{\mu_g}$ (red dashed line, Figure A.1F). The Poisson model, which is parameter-free, provides a tight lower bound of the CV for lowly expressed genes – i.e. the actual CV values for lowly expressed genes are equal to or higher than the Poisson CV. However, at high expression values we observed that the model significantly underestimates the minimum CV in the data. More specifically, the CV of genes in the data appears to plateau at high mean expression, whereas the Poisson model predicts a square root decrease.

What accounts for the over-dispersion in the data at high mean expression values compared to the Poisson model? According to the Poisson model, which treats genes independent of each other, the total number of transcript counts per cell (N_{tot}) is a sum of independent Poisson random variables, and therefore is a Poisson random variable itself. However, this is not supported by our data as the variance of N_{tot} is approximately 389 times its mean in the larval droplet data. This over-dispersion of total number of transcripts per cell for highly expressed genes could be caused by many factors. Some may be biological like cell size and cell state. However, many others may be technical factors such as variations in lysis and RT efficiency, number of captured oligonucleotides or extent of RNA degradation between droplets.

Based on this hypothesis, we made a simple modification to the Poisson model by positing that the sampling rate of a gene in a given cell depends on its relative library size η. We hypothesized that,

$$\eta = \frac{N_{tot,i}}{\mathbb{E}N_{tot,i}} \qquad (2)$$

where $N_{tot,i}$ is the total number of molecules in cell *i* and $\mathbb{E}N_{tot,i}$ is its expectation across all cells. We note that $\mathbb{E}\eta = 1$. We found that a Gamma distribution with mean fixed at 1 provided an excellent fit for the empirical distribution of η in all of our droplet datasets (e.g. Figure A.1E). Given the empirical distribution of η, we used the R package MASS to estimate the scale factor α for the Gamma distribution. This leads to a model where every gene is sampled from a Poisson distribution with its rate being a random variable following a Gamma distribution (our parameterization of η makes this a single parameter model).

Fortunately, this Gamma-Poisson results in a closed form solution, wherein every gene follows a negative binomial distribution, $X_g \sim NB(r, p_g)$. Here, *r* and *p* represent the canonical parameters of the negative binomial distribution, the number of failures(*r*) and the success probability(*p*), which follow the relations,

$$r = \alpha, p = \frac{\mu_g}{\alpha + \mu_g} \qquad (3)$$

Using standard properties of the negative binomial distribution, we compute the CV-mean relationship as,

$$CV^2{}_g = \frac{1}{\mu_g} + \frac{1}{\alpha} \qquad (4)$$

Figure A.1F graphs this relationship (solid magenta line). As shown, for lowly expressed genes, we expect $\frac{1}{\mu_g} \gg \frac{1}{\alpha}$, and the curve reduces to the Poisson regime. This suggests that for lowly expressed genes, the variation is dominated by the Poisson fluctuations. However, for highly expressed genes $\mu_g \gg \alpha$, the model reduces to $CV_g = 1/\sqrt{\alpha}$, explaining the saturation observed in the data. This suggests that for highly expressed genes the variation is dominated by various technical factors (contributing to library size differences) that result in more over-dispersion than predicted by the Poisson model. Importantly, through this simple modification, we were able to provide excellent estimates for the lower bound in CV across the full range of expression values. We ranked the genes based on their distance from this null curve in log-space, i.e. $\log\left(\frac{CV_{observed}}{CV_{NB}}\right)$. We used the shape of the distribution of this quantity to estimate a cutoff value (0.3), above which genes were considered highly variable.

We note here that while equation (4) models the baseline CV-mean relationships in transcript counts for UMI based data as a tight lower bound (Figure A.1F), it significantly underestimates the CV-mean relationship observed in SMART-seq2 data (Figure A.4D and A.4E). Here, a related but more flexible model appears to perform better at capturing the mean-CV behavior (Brennecke et al., 2013),

$$CV^2{}_g = \frac{\alpha_1}{\mu_g} + \frac{1}{\alpha_2} \qquad (5)$$

This model *(Brennecke et al.*) (Figure A.4E, yellow line I $\alpha_1 = 205, \alpha_2 = 2.9$) better approximates the CV-mean relationship observed in SMART-seq2 data. This behavior is not peculiar to the SMART-Seq2 data presented in the paper. We reran the Poisson Gamma and the *Brennecke et al.* model on the read counts data from Tasic, Menon et al. and found that our model underestimates CV-mean relationship (Figure A.4G). This overdispersion likely results from bias in the non-UMI Smart-Seq2 data, because the Poisson Model accurately predicts mean-CV baseline in Drop-seq data of retinal bipolar neurons, produced by an alternative UMI-based protocol (Shekhar et al., 2016) (Figure A.4H). We also verified this by analyzing read count data of larval droplet dataset prior to UMI collapse, which should “retain” amplification biases and found that the model performs worse with non-UMI based droplet dataset (Figure 4F).

Dimensionality reduction using PCA and Graph Clustering

Dimensionality reduction was performed using principal component analysis (PCA), and statistically significant PCs were identified using the Jackstraw function in Seurat (Chung and Storey, 2015). 36 significant PCs were identified for larval and 30 significant PCs for identified for adult data. The scores of cells along these significant PCs were used to build a *k*-nearest neighbor graph, and partition the cells into transcriptionally distinct clusters using the smart local moving community detection algorithm (Waltman and Eck, 2013) as implemented in the FindClusters function in Seurat. Subsequently, t-distributed stochastic neighbor embedding (tSNE) (Hinton and Maaten, 2008) was used to embed the cells based on statistically significant PCs, to visualize the graph clustering output on a 2D map. We note that the tSNE coordinates were computed independently of the cluster labels. All initial clusters were subjected to additional iterative clustering to discover additional heterogeneity within the initial clusters (Figure A.1G). We found two additional droplet clusters (Hb05 and Hb12) and 1 additional SMART-Seq2 cluster by iterative clustering. We also verified that cells did not segregate based on their experimental batch id (Figure A.4C and Figure A.5C) by observing the contribution of each experimental batch to every cluster in both adult and larval datasets. Downsampling experiments and assessment of cluster purity and entropy of the downsampled clusters were performed as described previously (Shekhar et al., 2016). Adjusted Rand Index (ARI) for cluster consistency was calculated as described previously (Cooper, 2010).

Marker genes discovery and quantification of their specificity and precision

Markers were nominated by performing a differential expression analysis between the cells in the cluster of interest and the rest of the cells in the dataset (Figure A.1L). Markers' specificity and precision were quantified using a statistical test based on the area under the precision-recall curve (AUCPR). AUCPR is a quantitative measure of the balance between recall (the sensitivity of marker gene detection within the cluster of interest) and precision (accuracy of the quantitative levels of gene as a predictor of the correct cell type). Markers found by our analysis were either "digital" (expressed only in the marked cluster) with AUCPR values> 0.8, or analog (expressed at a higher level in the marked cluster, but also detectable in other clusters) with AUCPR values between

0.6-0.8. We compared the AUCPR values of marker and non-marker genes at a range of expression values and show that the marker genes have significantly higher AUCPR values compared to non-marker genes (Figure A.2C). Markers with low AUCPR values belong to smaller clusters in which a small number of false positives in other clusters can significantly reduce the AUCPR value (Figure A.2C, right panel).

Comparison of cluster signatures between droplet and SS2 datasets

Independent analysis of SMART-seq2 data revealed 10 clusters (Figure 1.4A), fewer than in the droplet data (Figure 1.1C). We hypothesized that the lower sample size might have masked subtle transcriptional differences between closely related sub-types, causing the corresponding clusters to merge. We reasoned that a supervised classifier trained on the signatures of the droplet dataset might be able to resolve these merged clusters in the SS2 dataset. Therefore, to evaluate the correspondence between the droplet and SS2 clusters rigorously, we trained a multi-class random forest classifier on the droplet dataset. A random forest is an ensemble learning method that consists of multiple decision trees, each of which are trained on a randomly defined set of features (genes) (Breiman, 2001).

We composed a training set for the classifier by taking a sampling of cells from the 16 clusters from the droplet dataset. The number of training cells (N_k) from each cluster k was chosen such that N_k = min (500, $|cells_k|$*0.7), leading to the use of a maximum of 70% of the cells in each cluster for training. The remaining 30% of the cells in every cluster were used to test the classifier. In addition, the classifier was built on the most variable genes across both droplet and SS2 datasets. We trained the random forest using 1,000 trees on the training set with the R package randomForest. This trained classifier was then used to assign a cluster label for the remaining 30% of the data. We assigned a class label to each cell, but only if a minimum of 15% of trees in the forest converged onto a decision (given that there are 16 classes, 6.25% vote would constitute a majority). Otherwise, the cells were labeled unassigned. Cells in the test set were accurately mapped to their correct classes by the trained classifier at a median rate of 95% for every cluster as reflected by the diagonal structure of the confusion matrix (Figure A.4I).

This classifier was then used to predict the cluster labels of the cells in the SS2 dataset. It is important to note that the assignment is completely agnostic to the SS2 cluster label. After classifying each SS2 cell independently, we asked whether there was any correspondence between the SS2 clusters and the RF assignments to the droplet clusters. If greater than 70% of the cells of a single SS2 cluster mapped to single droplet clusters, that mapping was considered to be a 1:1 mapping. The Adjusted Rand Index (ARI, a measure of clustering consistency) between the RF assigned cluster labels and the SS2 labels was 0.45, significantly higher compared to the ARI between randomly permuted cluster labels and SS2 labels, which produced a maximum ARI value of 0.02 ($p < 0.001$).

Taken together, these results show that neuronal type-specific gene expression signatures that are important for cell type classification are robustly captured by sampling larger number of cells (with droplet) than a larger number of genes (with plate-based SS2).

Computation of entropy of cluster specific and extra SS2 genes

Compared to the droplet dataset, a host of extra genes in the SS2 dataset were expressed in a high proportion of cells but at lower expression level (Figures A.4E, A.4F). To understand the contribution of these “extra” genes in SS2, we identified all the genes that are expressed in moderate to high proportion of cells in the SS2 dataset but in a low proportion or not expressed in the droplet dataset (Figure A.4L, red). We then tested if these genes could serve as good cluster-specific markers or enabled better cell type classification.

To assess the ‘cluster-specificity’ of these genes, we computed their Shannon entropy, H_g:

$$H_g = \sum_1^n p(x) \ln p(x)$$

Where p(x) is the probability of the finding gene g in cluster x. For a single gene g, H_g equals 0 if the gene is expressed in a single cluster, but increases in value if the gene is expressed across multiple clusters.

We compared the distribution of the Shannon entropy (low H_g, measure of cluster-specificity) for these “extra” SS2 genes to its distribution among marker genes, which as expected exhibited low values of H_g. A large majority of the "extra" genes detected in SS2 dataset had higher entropy values, suggesting that they are uninformative for cell type classification (Figure A.4M).

Comparison of cluster signatures between larval and adult dataset

To systematically compare the larval and adult clusters, we used the expression of *gng8* to partition the adult clusters into "high" (average log (TPM + 1) > 2) and "low" (average log (TPM +1) < 2) *gng8*$^+$ cells. We used the "high" *gng8*$^+$ cells for further analysis (Figure A.5F) as the *gng8*-GFP transgenic line was used to capture cells in the larval dataset (Figure 2.1B). We trained a multi-class random forest classifier on our larval *gng8*-GFP droplet dataset as described above. We then used the RF model trained on the cluster signatures of the larval dataset to classify each adult cell into one of the larval habenular cluster labels, independent of the cell's adult cluster label. After classifying each adult cell independently, we asked whether there was any correspondence between the original clusters of the adult dataset and the RF-assigned clusters. This result was represented in the form of a confusion matrix as described in the SS2 cluster comparison section (Figure 2.5C). If greater than 70% of the cells of a single adult cluster mapped to a single larval cluster, that mapping was considered to be a 1:1 mapping. We verified that none of the adult cells mapped to larval cluster Hb16, which comprises olfactory placode cells that are labeled by the *gng8-GFP* transgenic line. These cells were not captured in the adult dataset as the adult dataset was generated by clean dissection of the habenula.

Functional experiments

Stimulation with electric shocks

For shock delivery, fish were placed in 85mm petri dishes over which a 10 milli-second 7-V shock was delivered every 30 seconds for 30 minutes. Current was delivered using alligator clips as in a previously described apparatus (Valente et al., 2012). Controls were placed beside the shocked fish and were also affixed with alligator clips but did not receive any shocks. Fish were then promptly fixed in 4% paraformaldehyde and subsequent RNA-FISH was performed for *cfos* expression.

These experiments were performed similarly in adults, except fish were kept in breeding cages during stimulation. After treatment with aversive stimuli, adult fish were rapidly killed by immersing in ice-cold water and their brains dissected immediately. Subsequent RNA-FISH was performed as described above to check for *cfos* expression in sub-populations within the habenula.

QUANTIFICATION AND STATISTICAL ANALYSIS

For experiments with larval habenula, cells were collected in three batches of SMART-seq2 data and one batch of 10X droplet data. For 10X experiment, 20 animals were pooled together in a single experiment. For SMART-seq2 experiments, batch 1 and batch 3 were composed of single animals whereas batch 2 was composed of 20 larval animals pooled together. For experiments with the adult habenula, cells were collected from age matched fish in two different experimental batches. Each batch was composed of 6 habenulae pooled together. For experiments involving RNA-FISH, each experiment was performed in batches of 20 fish. Approximately, 1000 larvae were imaged in total. For functional experiments in both larval and adult fish, experiments were performed in three independent batches. Within each batch, there were 20 treatment and 20 control animals for larva and 3 treatment and 3 control for adults.

All analysis related to RNA-seq was performed in R. Statistical methodologies and software used for performing various analysis in the paper are cited in appropriate STAR methods section. Differential expression of genes across clusters in the 10X and SMART-Seq2 experiments was evaluated using bimodal and binomial test as described before(Shekhar et al., 2016). Image registration analysis was performed in MATLAB and were parallelized using Slurm-based bash scripts.

REFERENCES

Agetsuma, M., Aizawa, H., Aoki, T., Nakayama, R., Takahoko, M., Goto, M., Sassa, T., Amo, R., Shiraki, T., Kawakami, K., *et al.* (2010). The habenula is crucial for experience-dependent modification of fear responses in zebrafish. Nat Neurosci *13*, 1354-1356.

Amo, R., Aizawa, H., Takahoko, M., Kobayashi, M., Takahashi, R., Aoki, T., and Okamoto, H. (2010). Identification of the zebrafish ventral habenula as a homolog of the mammalian lateral habenula. J Neurosci *30*, 1566-1574.

Amo, R., Fredes, F., Kinoshita, M., Aoki, R., Aizawa, H., Agetsuma, M., Aoki, T., Shiraki, T., Kakinuma, H., Matsuda, M., *et al.* (2014). The habenulo-raphe serotonergic circuit encodes an aversive expectation value essential for adaptive active avoidance of danger. Neuron *84*, 1034-1048.

Appelbaum, L., Wang, G.X., Maro, G.S., Mori, R., Tovin, A., Marin, W., Yokogawa, T., Kawakami, K., Smith, S.J., Gothilf, Y.*, et al.* (2009). Sleep-wake regulation and hypocretin-melatonin interaction in zebrafish. Proc Natl Acad Sci U S A *106*, 21942-21947.

Bayramli, X., Kocagoz, Y., Sakizli, U., and Fuss, S.H. (2017). Patterned Arrangements of Olfactory Receptor Gene Expression in Zebrafish are Established by Radial Movement of Specified Olfactory Sensory Neurons. Sci Rep *7*, 5572.

Beretta, C.A., Dross, N., Guiterrez-Triana, J.A., Ryu, S., and Carl, M. (2012). Habenula circuit development: past, present, and future. Front Neurosci-Switz *6*.

Bianco, I.H., and Wilson, S.W. (2009). The habenular nuclei: a conserved asymmetric relay station in the vertebrate brain. Philos Trans R Soc Lond B Biol Sci *364*, 1005-1020.

Biran, J., Palevitch, O., Ben-Dor, S., and Levavi-Sivan, B. (2012). Neurokinin Bs and neurokinin B receptors in zebrafish-potential role in controlling fish reproduction. Proc Natl Acad Sci U S A *109*, 10269-10274.

Breiman, L. (2001). Random forests. Mach Learn *45*, 5-32.

Brennecke, P., Anders, S., Kim, J.K., Kolodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C.*, et al.* (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods *10*, 1093-1095.

Broms, J., Antolin-Fontes, B., Tingstrom, A., and Ibanez-Tallon, I. (2015). Conserved expression of the GPR151 receptor in habenular axonal projections of vertebrates. J Comp Neurol *523*, 359-380.

Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M., Resch, J.M., McCarroll, S.A.*, et al.* (2017). A molecular census of arcuate hypothalamus and median eminence cell types. Nat Neurosci *20*, 484-496.

Chen, R., Wu, X., Jiang, L., and Zhang, Y. (2017). Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. Cell Rep *18*, 3227-3241.

Chou, M.Y., Amo, R., Kinoshita, M., Cherng, B.W., Shimazaki, H., Agetsuma, M., Shiraki, T., Aoki, T., Takahoko, M., Yamazaki, M., *et al.* (2016). Social conflict resolution regulated by two dorsal habenular subregions in zebrafish. Science *352*, 87-90.

Chung, N.C., and Storey, J.D. (2015). Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics *31*, 545-554.

Cooper, G.W.M.M.C. (2010). A Study of the Comparability of External Criteria for Hierarchical Cluster Analysis. Multivariate Behavioral Research, 441-458.

deCarvalho, T.N., Subedi, A., Rock, J., Harfe, B.D., Thisse, C., Thisse, B., Halpern, M.E., and Hong, E. (2014). Neurotransmitter map of the asymmetric dorsal habenular nuclei of zebrafish. Genesis *52*, 636-655.

Dreosti, E., Lopes, G., Kampff, A.R., and Wilson, S.W. (2015). Development of social behavior in young zebrafish. Front Neural Circuits *9*, 39.

Dreosti, E., Vendrell Llopis, N., Carl, M., Yaksi, E., and Wilson, S.W. (2014). Left-right asymmetry is required for the habenulae to respond to both visual and olfactory stimuli. Curr Biol *24*, 440-445.

Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M.V., Sudhof, T.C., and Quake, S.R. (2016). Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. Cell Rep *16*, 1126-1137.

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F., and Regev, A. (2016). Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science *353*, 925-928.

Hinton, G., and Maaten, L.v.d. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research 9, 2579-2605.

Hirai, H., Pang, Z., Bao, D.H., Miyazaki, T., Li, L.Y., Miura, E., Parris, J., Rong, Y.Q., Watanabe, M., Yuzaki, M., *et al.* (2005). Cbln1 is essential for synaptic integrity and plasticity in the cerebellum. Nature Neuroscience *8*, 1534-1541.

Hong, E., Santhakumar, K., Akitake, C.A., Ahn, S.J., Thisse, C., Thisse, B., Wyart, C., Mangin, J.M., and Halpern, M.E. (2013). Cholinergic left-right asymmetry in the habenulo-interpeduncular pathway. Proc Natl Acad Sci U S A *110*, 21171-21176.

Jefferis, G.S., Potter, C.J., Chan, A.M., Marin, E.C., Rohlfing, T., Maurer, C.R., Jr., and Luo, L. (2007). Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. Cell *128*, 1187-1203.

Jetti, S.K., Vendrell-Llopis, N., and Yaksi, E. (2014). Spontaneous activity governs olfactory representations in spatially organized habenular microcircuits. Curr Biol *24*, 434-439.

Kishimoto, N., Asakawa, K., Madelaine, R., Blader, P., Kawakami, K., and Sawamoto, K. (2013). Interhemispheric asymmetry of olfactory input-dependent neuronal specification in the adult brain. Nat Neurosci *16*, 884-888.

Kitahashi, T., Ogawa, S., and Parhar, I.S. (2009). Cloning and expression of kiss2 in the zebrafish and medaka. Endocrinology *150*, 821-831.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Kuan, Y.S., Yu, H.H., Moens, C.B., and Halpern, M.E. (2007). Neuropilin asymmetry mediates a left-right difference in habenular connectivity. Development *134*, 857-865.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., *et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Marquart, G.D., Tabor, K.M., Horstick, E.J., Brown, M., and Burgess, H.A. (2016). High precision registration between zebrafish brain atlases using symmetric diffeomorphic normalization.

Masland, R.H. (2001). The fundamental plan of the retina. Nature Neuroscience *4*, 877-886.

McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461-467.

Ngo, L., Haas, M., Qu, Z., Li, S.S., Zenker, J., Teng, K.S., Gunnersen, J.M., Breuss, M., Habgood, M., Keays, D.A.*, et al.* (2014). TUBB5 and its disease-associated mutations influence the terminal differentiation and dendritic spine densities of cerebral cortical neurons. Hum Mol Genet *23*, 5147-5158.

Okamoto, H. (2014). Neurobiology: sensory lateralization in the fish brain. Curr Biol *24*, R285-287.

Okamoto, H., and Aizawa, H. (2013). Fear and anxiety regulation by conserved affective circuits. Neuron *78*, 411-413.

Pandey, S., Shekhar, K., Regev, A., and Schier, A.F. (2018). Comprehensive Identification and Spatial Mapping of Habenular Neuronal Types Using Single-Cell RNA-Seq. Curr Biol *28*, 1052-1065 e1057.

Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods *10*, 1096-1098.

Pollen, A.A., Nowakowski, T.J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C.R., Shuga, J., Liu, S.J., Oldham, M.C., Diaz, A.*, et al.* (2015). Molecular identity of human outer radial glia during cortical development. Cell *163*, 55-67.

Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P.*, et al.* (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat Biotechnol *32*, 1053-1058.

Poulin, J.F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J.M., and Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. Nat Neurosci *19*, 1131-1141.

Randlett, O., Wee, C.L., Naumann, E.A., Nnaemeka, O., Schoppik, D., Fitzgerald, J.E., Portugues, R., Lacoste, A.M., Riegler, C., Engert, F.*, et al.* (2015). Whole-brain activity mapping onto a zebrafish brain atlas. Nat Methods *12*, 1039-1046.

Rohlfing, T., and Maurer, C.R., Jr. (2003). Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. IEEE Trans Inf Technol Biomed *7*, 16-25.

Ronneberger, O., Liu, K., Rath, M., Ruebeta, D., Mueller, T., Skibbe, H., Drayer, B., Schmidt, T., Filippi, A., Nitschke, R.*, et al.* (2012). ViBE-Z: a framework for 3D virtual colocalization analysis in zebrafish larval brains. Nat Methods *9*, 735-742.

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

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D.*, et al.* (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236-240.

Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M.*, et al.* (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell *166*, 1308-1323 e1330.

Shimada, Y., Hirano, M., Nishimura, Y., and Tanaka, T. (2012). A high-throughput fluorescence-based assay system for appetite-regulating gene and drug screening. PLoS One *7*, e52549.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T.*, et al.* (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci *19*, 335-346.

Valente, A., Huang, K.H., Portugues, R., and Engert, F. (2012). Ontogeny of classical and operant learning behaviors in zebrafish. Learn Mem *19*, 170-177.

Waltman, L., and Eck, N.J.v. (2013).

Wong, M.K., Sze, K.H., Chen, T., Cho, C.K., Law, H.C., Chu, I.K., and Wong, A.O. (2013). Goldfish spexin: solution structure and novel function as a satiety factor in feeding control. Am J Physiol Endocrinol Metab *305*, E348-366.

Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C.*, et al.* (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138-1142.

Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. Nat Rev Neurosci *18*, 530-546.

Zheng, G.X., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M.*, et al.* (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat Biotechnol *34*, 303-311.

Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J.*, et al.* (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049.

CHAPTER 3: Functional Investigation of Ventral Habenular Cell Types and Cluster-specific Marker Genes

3.1 PREFACE

I designed, performed, analyzed and interpreted the experiments with input from Alexander F. Schier. William Joo built the behavior rigs used for performing behavioral screening with the mutants. Summer Thyme generated the pipeline for extraction of behavioral features from behavior videos. Kristian Herrera helped with the calcium imaging experiments. Alexander F. Schier supported the project. Identification of ventral cell population responsive noxious electric shocks was published in Current Biology.

3.2 ABSTRACT

Brain function depends on the complex network of specialized neuronal types that give rise to system-level behavioral characteristics of animals. Single cell technologies allow comprehensive mapping of the repertoire of cell types in the brain in unprecedented detail. However, a key next step in understanding the role of the cell types as well as the cell type specific marker genes in generating brain function is to map an additional layer of functional properties onto the molecular descriptors from single-cell RNAseq. In this chapter, I present two distinct ways in which I map functionality onto molecular descriptors of the habenula. In the first, approach I leverage the spatial map of the

habenula and immediate early genes to localize cell populations responsible for generating responses to inescapable aversive environmental stimuli. Using *cfos* as a neural activity indicator, I discovered a population in the ventrolateral habenula that robustly responds to noxious electric shocks. In the second part, I employ CRISPR-Cas9 to generate knockouts of genes that are specific to sub-clusters within the ventral habenula. All of the four mutants generated in our study had various locomotor defects along with changes in baseline brain activity. Particularly, notable among these genes is *aoc1*, a histamine inactivating enzyme that is specific to the sub-clusters of the ventral habenula and may have an important role in regulating sleep. A detailed analysis of brain activity and behavioral patterns of mutant animals suggests that all the genes expressed within the ventral habenula may have important roles in regulating locomotion.

3.3 INTRODUCTION

Single cell RNA-seq deconvolutes the quagmire of neurons in the brain into specific cell types by proximity of their molecular signatures. These molecular signatures form the core molecular identity of said neurons. However, a mechanistic understanding of the brain at a functional level requires that these cell types and marker genes be associated with specific functional roles with respect to generating animal behaviors in both normal and diseased states. In short, the single cell RNAseq analysis of the brain can be a genetic scaffold onto which other cellular phenotypes can be overlaid.

To extend our single cell analysis of the habenula into functional regimes, I took two different approaches. In the first approach, I used spatial localization of cell types to link molecularly described cell types to their anatomical positions and assign them to specific functional roles using the expression of immediate early gene *cfos*. This approach was similar to the multiplexed in situ hybridization approaches that were used by recent studies to characterize different parts of the mammalian brain (Moffitt et al., 2018; Shah et al., 2016) but limited by the fact that the multiplexing was performed computationally.

In the second approach, I investigated the role of specific marker genes expressed in ventral habenular clusters. The ventral habenula, the homolog of the mammalian lateral habenula, plays important roles in sleep, reward learning and coping with aversive environmental states (Aizawa et al., 2013a; Andalman et al., 2019; Wang et al., 2017). While the role of the general ventral habenular region has been explored extensively in these contexts, it is unclear whether the ventral habenular cell type specific markers play important roles in mediating any of these behaviors. Therefore, this region of the habenula represented a perfect opportunity to connect gene signatures specific to habenular cell types to behavioral functions mediated by those types. In this chapter, I will discuss efforts to overlay functional properties into ventral habenular clusters using these two complementary approaches.

3.4 RESULTS

3.4.1 Computational Image Registration Generates a Spatial Map of Neuronal Types in the Habenula

The study of brain function ultimately requires linking molecularly distinct cell types to measures of function. Therefore, it is critical to place cell types identified by analysis of global gene expression data obtained using scRNA-seq to its native spatial context by using top cluster-specific markers. To explore the localization of neuronal types within the habenula, I created a consolidated spatial map of neuronal type specific markers in the habenula (Figure 3.1). I used computational image registration to morph RNA-FISH signals of many cluster-specific marker genes across multiple zebrafish larvae onto a single reference based on the total ERK (tERK) antibody stain (Jefferis et al., 2007; Randlett et al., 2015; Rohlfing and Maurer, 2003) (Figure A.3) resulting in a reference spatial map of the habenula.

My map demonstrates that a majority of habenular types were regionalized either along the dorso-ventral, medio-lateral or left-right axis, suggesting that habenular types can be defined not only by their transcriptomes but also by distinct spatial positions (Figure 3.1B). Eleven types were located more dorsally in the habenula; 8 of them seem to be regionalized: Hb01/Hb02 were right enriched or exclusive, Hb07/Hb08/Hb09/Hb10 were left-enriched or exclusive, Hb04 and Hb06 was L-R symmetric in the dorsal habenula. Four types were located more ventrally, of which Hb11 and Hb12 were posterior ventral populations, Hb15 occupied a large portion of the ventral habenula and Hb13 formed a germinal zone along the medial region of the ventral habenula. In summary, the fluorescent *in situ* atlas provides the first spatial map of the 15 known and novel neuronal types in the zebrafish habenula.

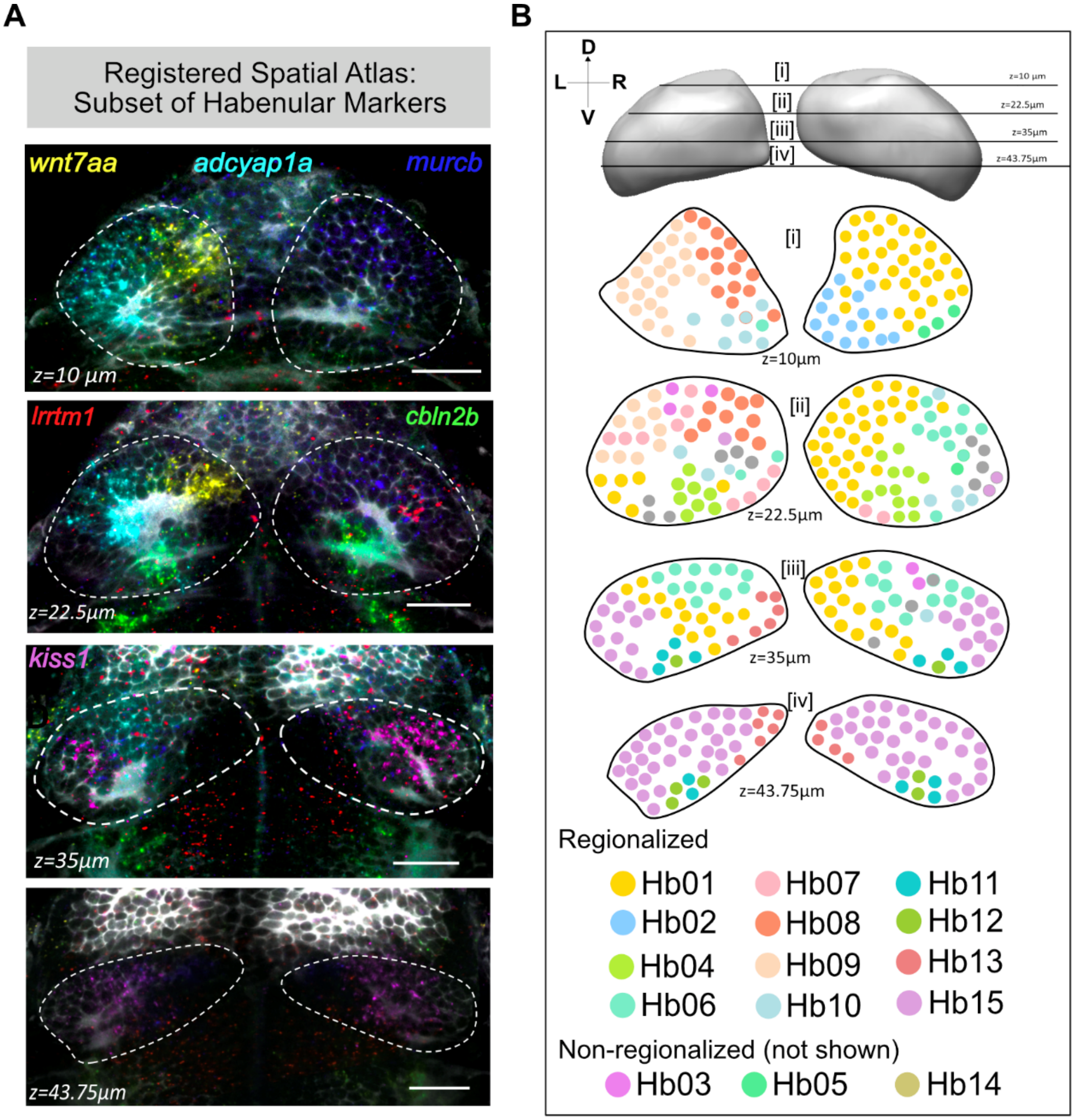


Figure 3.1. Spatial Atlas of the Zebrafish Habenula

(A) Slices through the registered reference habenula simultaneously showing six marker genes that are expressed in a regionalized pattern: *wnt7aa* (La_Hb08), *adcyap1a* (La_Hb07), *cbln2b* (La_Hb04), *murcb* (La_Hb01), *lrrtm1* (La_Hb06), *gpr139* (La_Hb15).

Figure 3.1 (Continued). (B) Schematic of representative transverse slices through the habenula displaying rough spatial co-ordinates of previously described as well as new neuronal types found by single-cell analysis. Cells are color-coded based on their identity in the t-SNE plot (see Figure1C). Depth is indicated by the z slice in microns. The sectioning extends from z = 0μm (Dorsal) and z = 75μm (Ventral). Only regionalized markers are represented. Schematic is a simplified representation of an accompanying stack of registered habenular markers overlaid onto one another [see Movie S1]. Scale bars indicate 50 μm.

3.4.2 Functional Assignment onto Molecularly Defined Neuronal Types

Next, I leveraged the spatial atlas generated by registration of in situ data to assign functional properties to molecular subtypes. I began by exploring if and how functional responses of the habenula were distributed among the transcriptionally distinct neuronal types. To this end, I performed a proof-of-principle experiment to determine if we could use neuronal type specific markers and the reference habenular atlas (Figure 3.1A) to map responses to specific stimuli among the molecularly defined neuronal types. One of the key roles of the zebrafish habenula is to generate behavioral responses to inescapable aversive environmental stimuli. However, it is unclear which neuronal subtype within the habenula mediates this behavior. I exposed fish to noxious inescapable electric shocks and assessed upregulation of the immediate early marker gene *cfos* in specific neuronal types. I observed localized upregulation of *cfos* expression in specific sub-regions within the habenula in shocked animals (Figure 3.2A). I registered the *cfos* RNA-FISH signals onto our reference habenular map(Figure 3.3A), and detected overlap within the domains of expression of left habenular Hb09 (*adcyap1a*$^+$), posterior habenular Hb04 (*cbln2b*$^+$) and largely within the ventrolateral population (*mprip*$^+$*)* (Figure 3.2)*.* Double *in situ* hybridization for *mprip* and *cfos* showed a regional overlap between the *mprip*$^+$ and *cfos*$^+$ cells in both larva (Figure 3.3C) and adult (Figure 3.3D). Together, these results show that neuronal ensembles in the ventrolateral habenula comprise a large proportion of cells responsive to aversive electric shocks and demonstrate the utility of the habenular atlas to link functional responses to molecularly defined neuronal types.

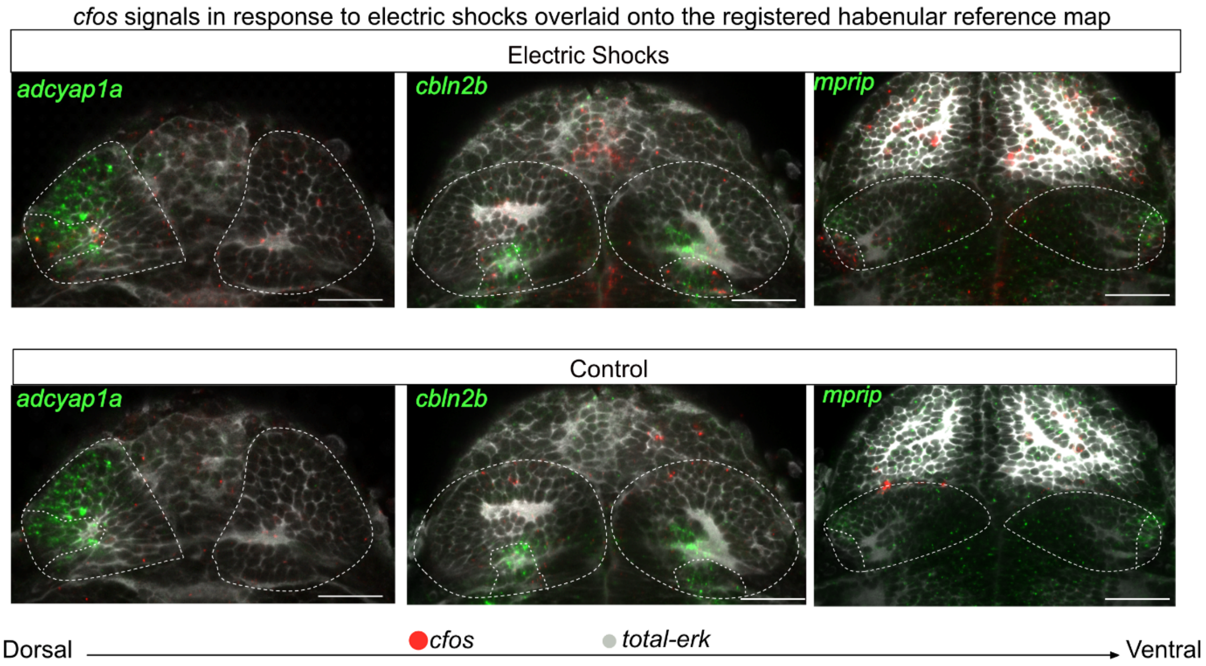


Figure 3.2: Mapping of *cfos* Responses onto Neuronal Types Found by scRNA-seq

Registration of *cfos* expression in response to aversive electric shocks onto molecular atlas generated by habenula registration shows major co-regionalization with markers for three clusters found by single cell dataset: Hb09 (*adcyap1a*), Hb04 (*cbln2b*) and ventrolateral cluster (*mprip*)**.**

These results suggest that the cell type along ventro-lateral habenula defined by the expression of *mprip* is important for responses to aversive environmental stimuli.

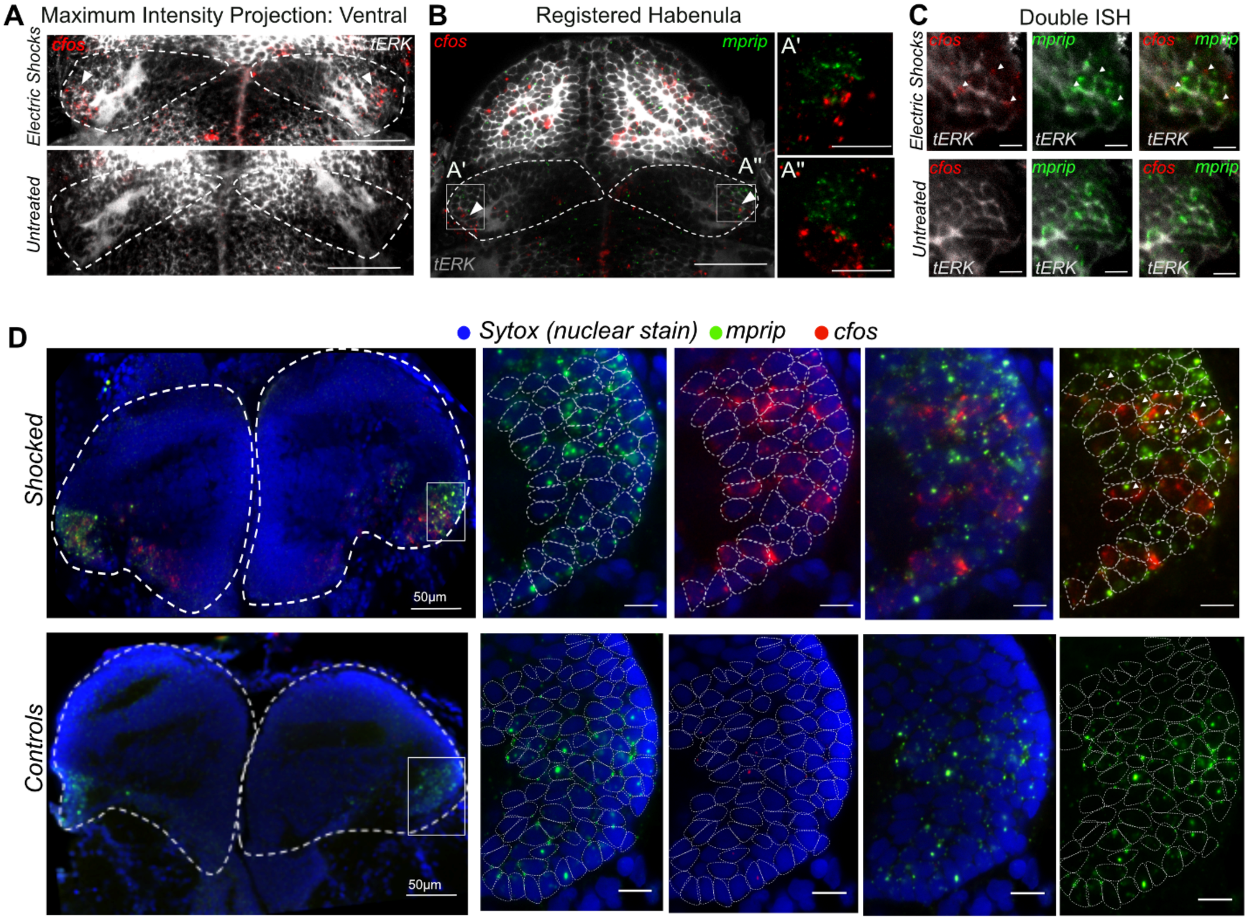


Figure 3.3. Noxious Electric Shocks Activate a Sub-Population of Neurons in the Ventro-lateral Habenula Labeled by *mprip.*

(A) ISH analysis of *cfos* expression in the habenula 30 minutes after exposure to electric shocks.

(B) Registration of *cfos* signals to habenular molecular atlas reveals co-regionalization with the *mprip*$^{+}$ ventrolateral population [Scale bars represent 50 µm].

(C) Double in situ hybridization for *c-fos* and *mprip*, (marker for ventrolateral neuronal type) showing a co-localization of *cfos*$^{+}$ and *mprip*$^{+}$ domains in the larval habenula in response to electric shocks [Scale bars represent 10 µm].

(D) Double in situ hybridization of *cfos* and *mprip* showing the conservation of electric shocks-induced *cfos* responses in *mprip*$^{+}$ ventro-lateral neuronal type in the adult habenula. Nuclei borders are demarcated in the zoomed in panels on the right using dotted circles [Scale Bar represents 10µm unless otherwise stated].

3.4.3 Generation of CRISPR knockouts for specific marker genes of ventral habenular clusters.

In a concurrent and complementary approach to understand the functional relevance of habenular cell types, I also interrogated the role of cell type specific marker genes in generating habenular function. To this end, I generated CRISPR knockouts (see Appendix for details on DNA lesions) for marker genes of the ventral habenular clusters. I chose four genes based on their restricted expression pattern among the ventral habenular clusters and large exclusion throughout the rest of the brain (Figure 3.4). More specifically, *aoc1* and *gpr139* were limited in their expression to Hb15 whereas *nts* and *ntsr1* are limited in their expression to Hb11/Hb15 and Hb11 respectively (Figure 3.4A).

To identify the neurobiological roles of these marker genes, I performed a phenotypic analysis of brain activity, morphology and baseline behavior. I assessed baseline motion over multiple day-night cycles and responses to sensory stimulation such as taps and dark flashes (Burgess and Granato, 2007; Kokel et al., 2010; Rihel et al., 2010). All phenotypic assays compared homozygous mutants to heterozygous and wild type siblings coming from the same clutch. In the following section, I describe the brain activity and behavioral phenotypes of the animals harboring mutations for these marker genes.

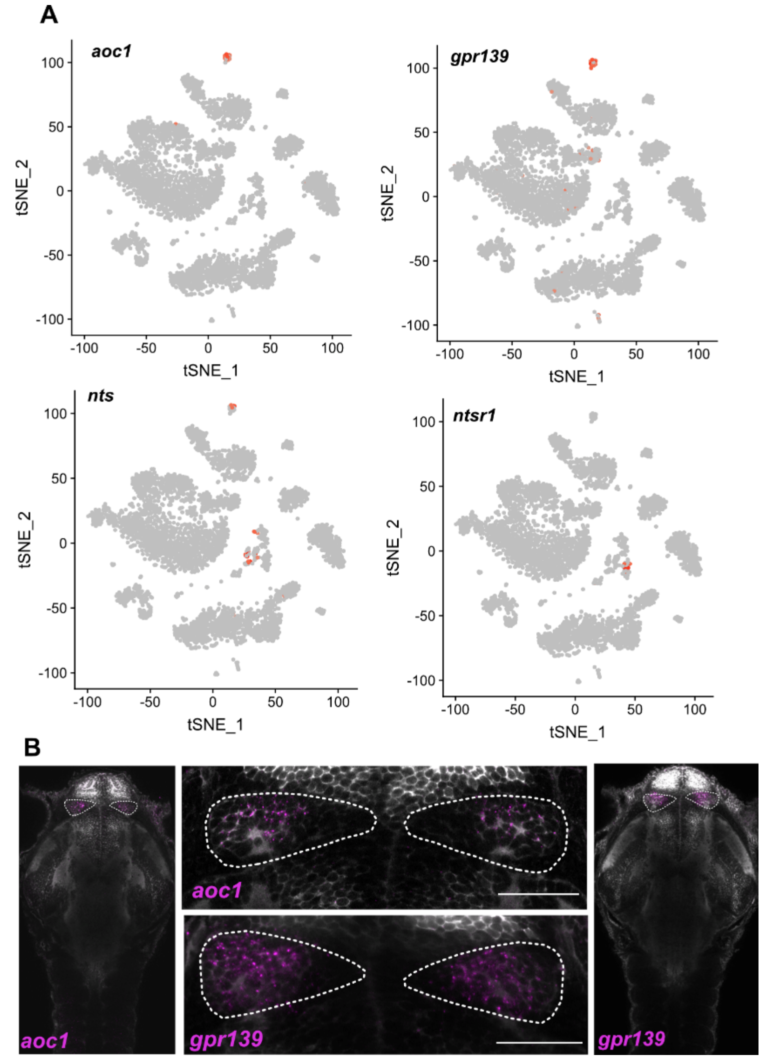


Figure 3.4. Expression Patterns of the Genes Knocked out from Ventral Habenular Clusters.

(A) Feature Plots showing the expression patterns of selected ventral habenula-specific genes that were used to interrogate the role of cell type specific genes in generating habenular function.

(B) In situ hybridization showing expression of aoc1 and gpr139 within the habenula (middle panel) and across the entire brain (left and right panels)

3.4.3.1 Gpr139, an orphan G-protein coupled receptor

Gpr139 is an orphan Gq-coupled receptor that is activated by essential amino acids such as L-tryptophan and L-phenylalanine (Liu et al., 2015). Recently, *gpr139* was also shown to be activated by endogenous peptides such as adrenocorticotrophic hormone (ACTH), melanocyte stimulating hormone (MSH) (Nohr et al., 2017). While the physiological function of *gpr139* remains elusive, it is highly associated with neuropsychiatric diseases such as depression (Dunn et al., 2016). The domain of expression of *gpr139* in larval zebrafish is limited to the ventral habenula (Figure 3.4A and B) and is excluded from the rest of the brain. This specificity in expression and its involvement in such neuropsychiatric disorders makes it a good candidate gene for exploring the functional relevance of habenular cell type specific genes. Given this expression pattern, it might have important roles in a number of habenula mediated behavioral states such as sleep, stress, anxiety and addiction. Concurrent to this, a recent study has reported that an intra-habenular activation of *gpr139* leads to decrease in alcohol dependence in rats (Kononoff et al., 2018). Therefore, we created a loss of function mutation of *gpr139* to investigate its role in generating ventral habenular function.

Using baseline brain activity measurement, we observed specific increases in brain activity in the *gpr139*-/- zebrafish larvae (Figure 3.5B). Activity differences were limited to the right ventral habenula and to the medial pallium. Furthermore, zebrafish larvae with mutations in *gpr139* displayed a slight increase in locomotor activity during the day, accompanied by an increase in the number of bouts (Figure 3.5C and D) and average bout speed (Figure 3.5L) during the day. When exposed to varying tap and strengths during the day and night to measure their arousal threshold, *gpr139-/-* larvae displayed no difference in the response probability at a range of tap strengths. On the other hand, when exposed to varying levels of dark flashes, *gpr139-/-* larvae showed a reduced response probability to dark flashes of lower intensity (Figure 3.5H).

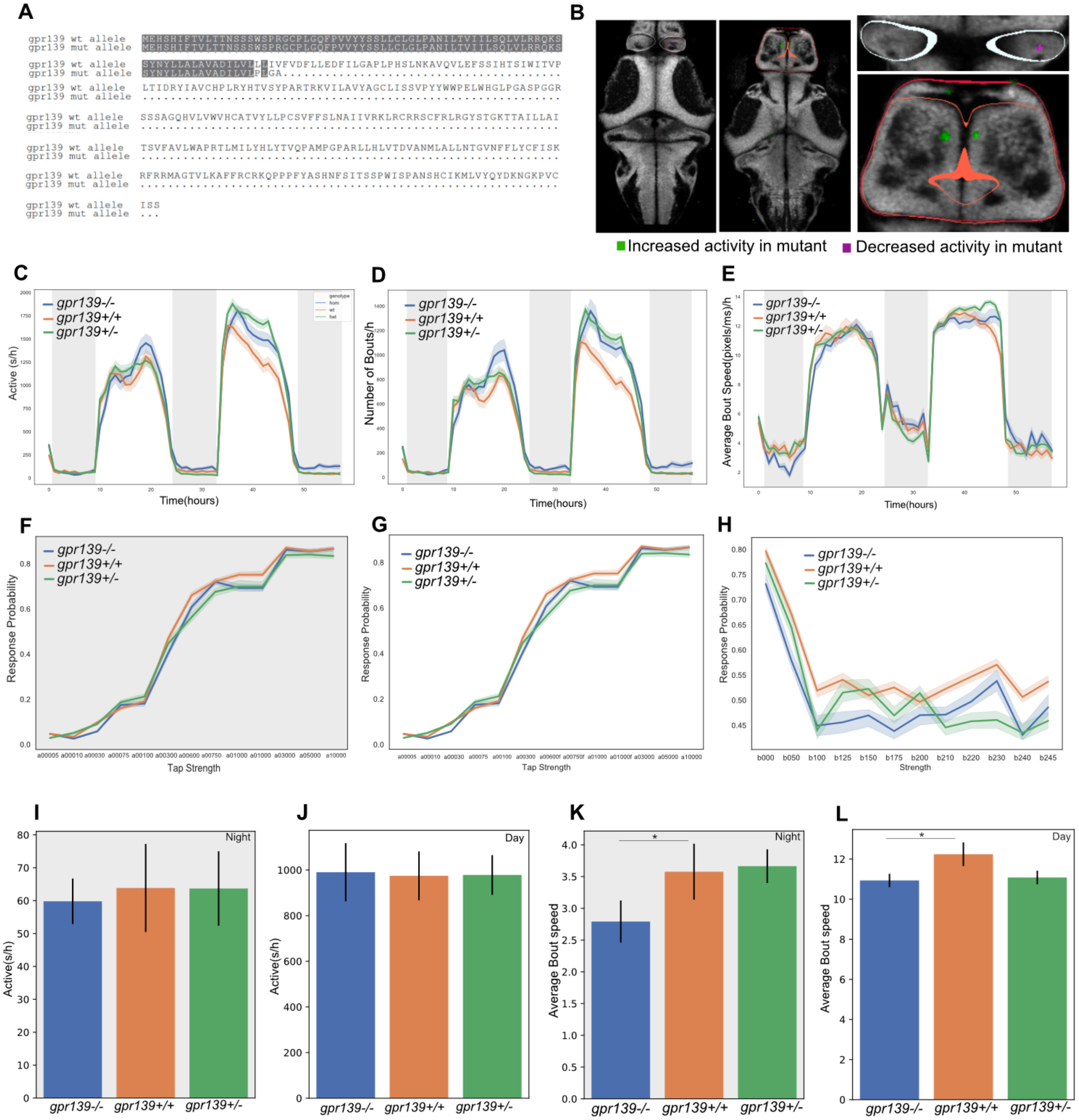


Figure 3.5. Loss of gpr139 decreases locomotor activity during the day.

(A) Sequences of wild type and mutant zebrafish gpr139 proteins.

Figure 3.5 (Continued). (B) Brain activity phenotypes representing baseline differences in brain activity between mutant and wild type fish.

(C-E) *Gpr139*-/- and *gpr139*+/- animals were slightly more active during the day and displayed increased number of bouts during the day but had the same average bout speed.

(F-G) *Gpr139*-/- animals displayed no difference in response probability to taps and dark flashes both during the day and night.

(I-J) *Gpr139*-/- animals displayed no significant change in activity during the day and night.

(K-L) *Gpr139*-/- animals displayed a significant decrease in the average bout speed compared to wild type siblings both during the day and at night.

3.4.3.2 Diamine oxidase (aoc1/dao), an enzyme that deaminates histamine.

Amine oxidase copper containing 1 or diamine oxidase (*aoc1/dao*) is a gene that encodes a metal binding secreted glycoprotein that oxidatively deaminates histamine, putrescine and other related compounds. Diamine oxidase is thought to be absent in the mammalian brain but present in the brain of lower vertebrates (Almeida and Beaven, 1981). In the zebrafish brain, the expression of *aoc1* is limited to the ventral habenular nucleus (Figure 3.4).

Pharmacological studies in both mammals and zebrafish have suggested that histamine plays an important role in promoting arousal (Thakkar, 2011). Studies in zebrafish have concluded that animals containing predicted null mutations in histamine receptors lack robust sleep/wake phenotypes (Chen et al., 2017). Similarly, zebrafish exhibit a dose-dependent decrease in locomotor activity when treated with histamine receptor antagonists (Sundvik et al., 2011). Histamine is synthesized by histidine decarboxylase (Hdc) which is expressed in the tuberomammillary nucleus (TMN) of the hypothalamus in mammals and in the ventrocaudal hypothalamus in the zebrafish brain (Sundvik and

Panula, 2012), Histamine can be inactivated in two distinct ways: 1) oxidative deamination by diamine oxidase (dao) which is thought to remove extracellular histamine (Schwelberger, 2004b). 2) methylation of the imidazole ring, catalyzed by histamine N-methyltransferase (HMT) which is thought to remove cellular histamine (Schwelberger, 2004a).

It has long been suspected that the habenula plays an important role in the regulation of sleep. Till date, this habenular regulation of sleep has been shown to occur via its regulation of downstream serotonin and dopaminergic targets (Aizawa et al., 2013b). However, the role of habenula in regulating sleep via effects exerted on the histaminergic center of the brain has been largely ignored. The selective expression of *aoc1* in the cell types in the ventral habenular nucleus might represent a parallel pathway via which habenula might exert its effects on sleep. Indeed, direct habenular projections to histaminergic systems have been found in lampreys in addition to the projections to the dopamine and serotonin centers (Stephenson-Jones et al., 2012).

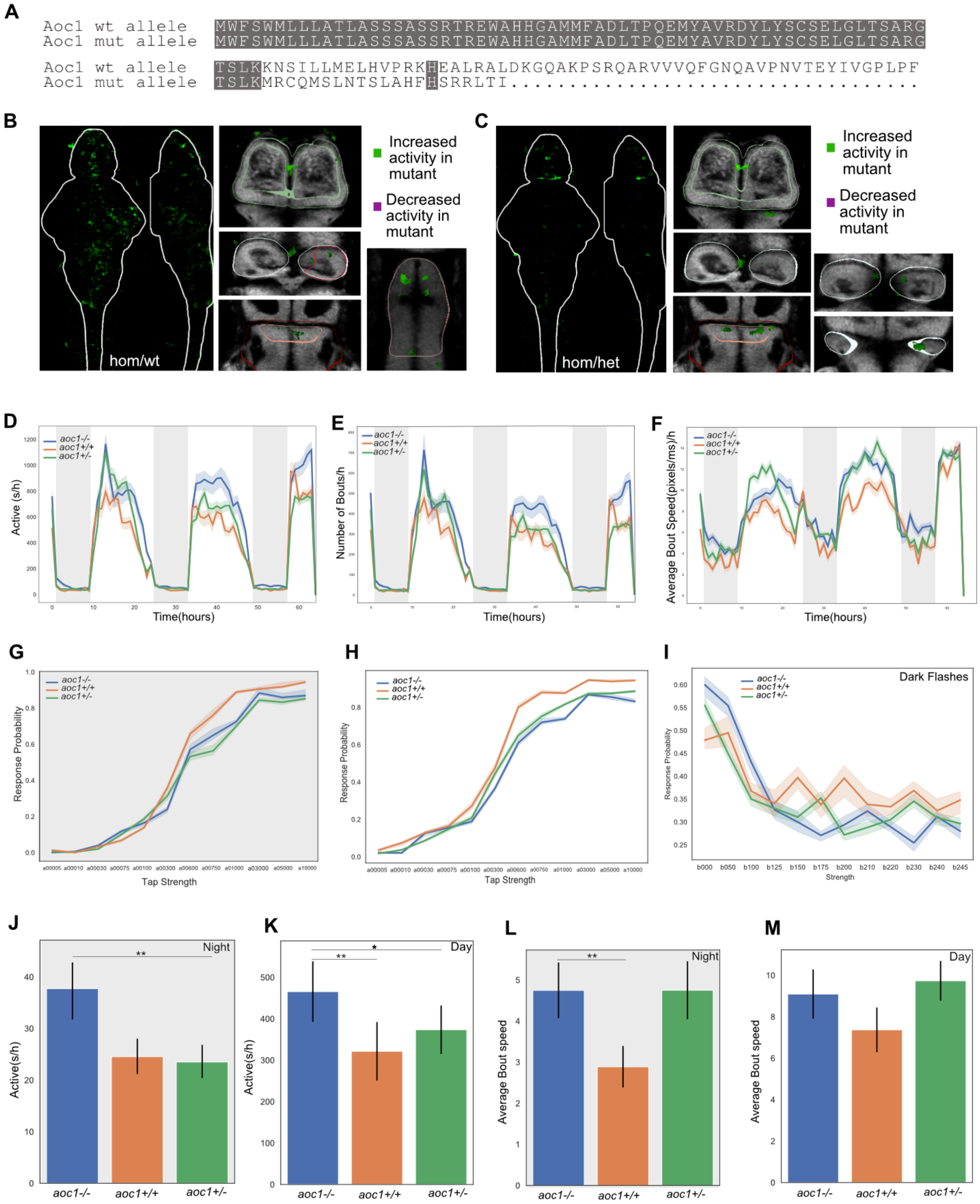


Figure 3.6. Loss of *aoc1* increases locomotor activity.

(A) Sequences of wild type and mutant zebrafish gpr139 proteins.

Figure 3.6 (Continued). (B) Brain activity phenotypes representing baseline differences in brain activity between *aoc1-/-* and *aoc+/+* fish.

(C) Brain activity phenotypes representing baseline differences in brain activity between *aoc1-/-* and *aoc+/-* fish.

(D-F) *aoc1*-/- animals were more active during the day and displayed more number of bouts and higher average bout speed during the day.

(G-I) *aoc1-/-* animals displayed slightly increased response probability to taps and dark flashes both during the day and night.

(J-K) *aoc1-/-* animals are more active during the day and night compared to wild type siblings.

(L-M) *aoc1*-/- animals displayed an increase in average bout speed during the night but not during the day.

Zebrafish larvae harboring mutations for *aoc1* also displayed specific increases in the brain activity in the pallium region of the forebrain and in the ventral habenula (Figure 3.6B and C). This change in brain activity was accompanied by an increase in locomotor activity during the day. The increase in daytime activity was caused by increase in the number of bouts. The mutants also displayed a significant increase in average bout speed during the night (Figure 3.6F, L, M). Interestingly, when exposed to varying tap and strengths during the day and night to measure their arousal threshold, *aoc1-/-* larvae displayed a small but significant decrease in the response probability at higher ranges of tap strengths suggesting that they may have an increased arousal threshold. However, there appeared to be no change in the response probability to dark flashes at all intensities (Figure 3.5H).

3.4.3.3 Neurotensin and Neurotensin receptor

Neurotensin (*nts*) is a neuromodulator of dopamine transmission and of anterior pituitary hormone secretion and is known to exert potent hypothermic and analgesic effects in the brain. In the zebrafish habenula, both neurotensin and its receptor were expressed in a small group of neurons in a cluster belonging to the ventral habenula (Figure 3.4A). However, their expression, unlike *aoc1* and *gpr139*, is not limited to the ventral habenula but is also seen in the hypothalamus and some parts of the forebrain (data not shown).

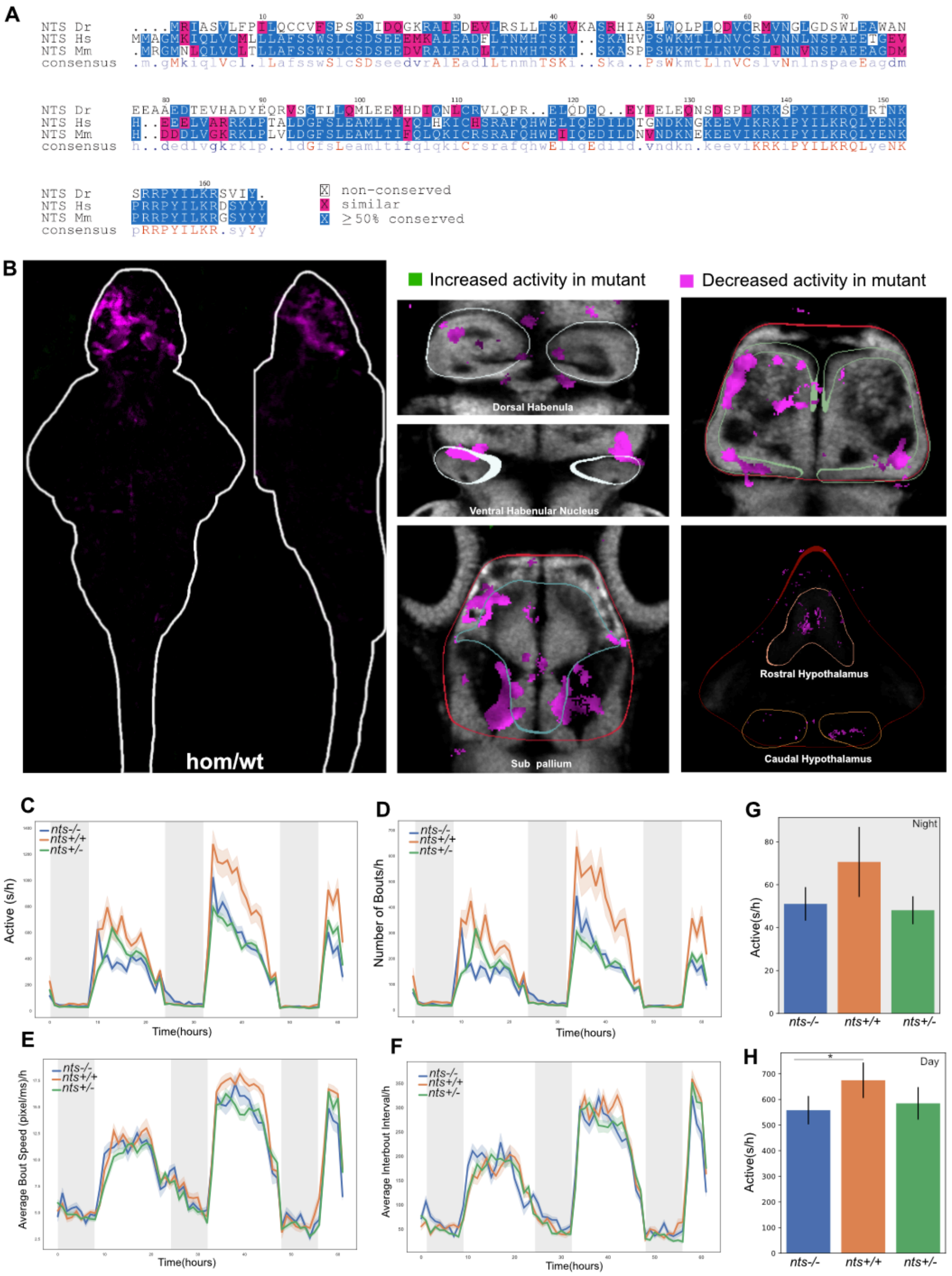


Figure 3.7. Loss of nts decreases daytime locomotor activity.

(A) Alignment of zebrafish nts sequences with mouse and human nts protein sequence.

(B-C) Brain activity phenotypes representing baseline differences in brain activity between nts-/- and nts +/+ fish.

(C-F) nts-/- displayed decreases in daytime activity (C) with corresponding increases in number of bouts (D) but no change in average bout speed or interbout interval (F).

(G-H) *nts*-/- animals are less active during the day and night.

Nts-/- zebrafish larvae displayed significant decreases in the brain activity in both the pallium and subpallium region of the forebrain (Figure 3.7B). In addition to that, they also displayed a strong decrease in activity in the ventral habenula and various parts of the rostral and caudal hypothalamus. This change in brain activity was accompanied by a decrease in locomotor activity during the day (Figure 3.7C and D). However, no changes were observed in the average bout speed or interbout interval (Figure 3.7E and F). *Ntsr1*-/- larvae in the other hand, displayed only mild increases in brain activity in the various regions of the hindbrain (Figure 3.8B and C). Contrary to *nts*-/- phenotypes, *ntsr1* mutants displayed no significant changes in locomotor activity during the day but displayed reduced locomotor activity at night (Figure 3.8H). They also displayed a significant increase in the average bout speed (Figure 3.8D and F).

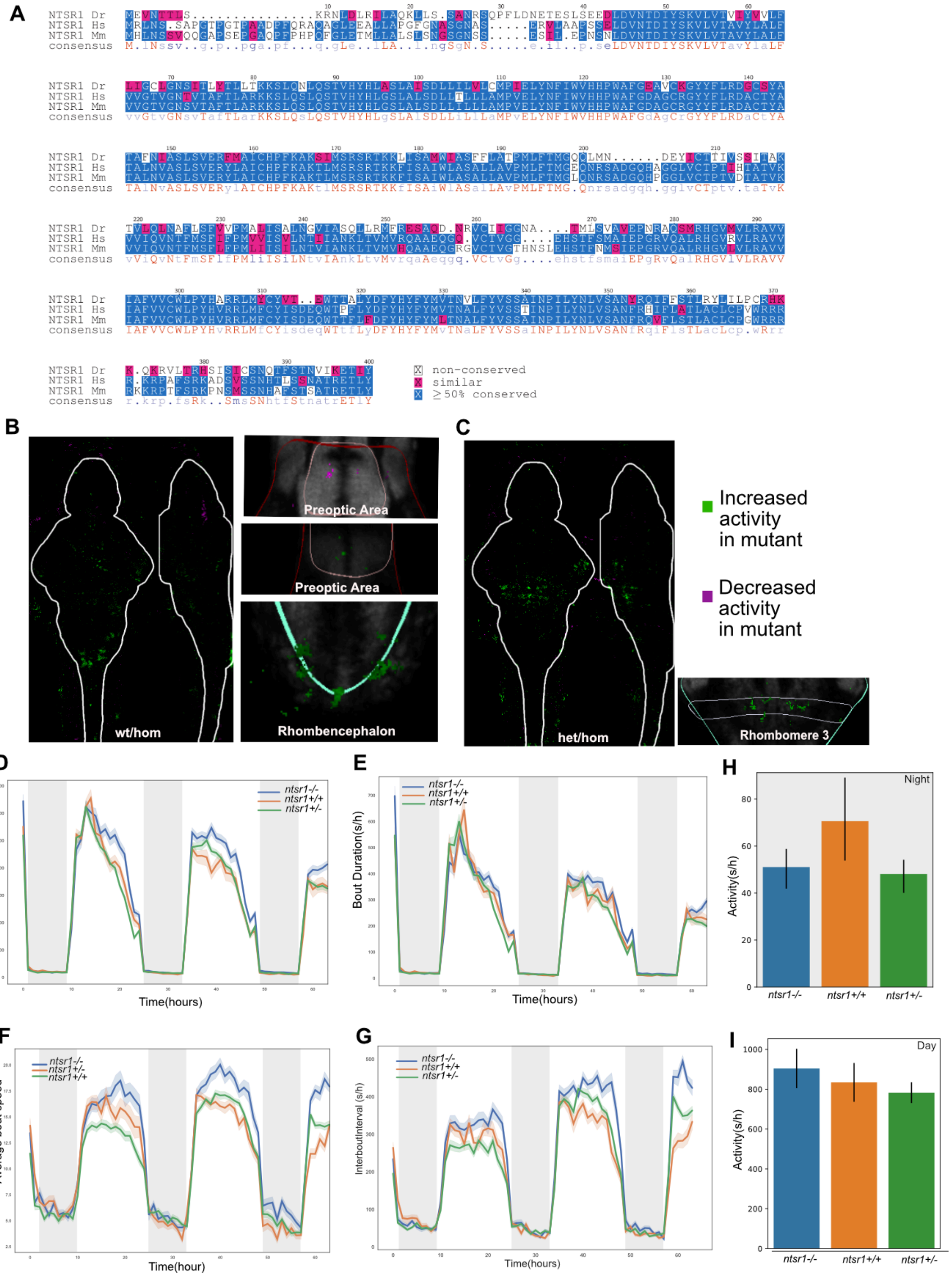


Figure 3.8: Loss of ntsr1 increases daytime locomotor activity.

(A) Alignment of zebrafish ntsr1 sequences with mouse and human ntsr1 protein sequence.

(B-C) Brain activity phenotypes representing baseline differences in brain activity between *ntsr1*-/- and *ntsr1*+/+ fish(B) and *ntsr1*-/- and *ntsr1*+/+ fish(C)

(D-F) ntsr1-/- displayed no change in daytime activity (D, I) with corresponding increases in bout speed (F) and interbout interval during the day(G). ntsr1-/- animals displayed no change in bout duration.

(H-I) *ntsr1*-/- animals were less active during the night but not during the day.

3.5 DISCUSSION

A number of studies have demonstrated that defining cell types in their natural habitat is an important next step for generating functionally relevant brain atlases (Moffitt et al., 2018; Shah et al., 2016). In this chapter, I presented how we utilized computational morphing techniques to generate a spatial atlas of the zebrafish habenula using markers obtained from single cell analysis. These analysis help add an additional layer of spatial information onto these molecular identities. Neuronal types were distinguished not just by their molecular profiles but by their regional expression patterns within the habenular tissue. We then used this atlas to identify a cell type in the ventral habenula that is important for responses to inescapable aversive stimuli. Since then, a number of studies have corroborated these findings to show that the ventrolateral cell type in the habenula is important for the responses to aversive environmental stimuli and that dynamics within these cell type domains are important for mediating a behavioral switch from active coping to passive coping strategies (Andalman et al., 2019). In future, it will be interesting to use these comprehensive and extensive single cell atlases with multiplexed in situ hybridization approaches such as MERFISH or seq-FISH to obtain a single cell resolution map of various fish brains (See Appendix Figure A.6 for preliminary data). In an approach similar to above, these in situ atlases could be interfaced with immediate early genes such as *cfos, egr1, npas4* in a high throughput manner for unbiased identification of cell types involved in a number of different behaviors. These approaches can provide a framework for mechanistic investigation of circuits that are involved in various behavioral paradigms.

Our work on the characterization of marker genes using knockouts represents a parallel approach to overlaying functional information onto single cell clusters. We chose four genes that are specifically expressed in clusters in the ventral habenula and to a large part were excluded from the rest of the brain and assayed both brain activity and behavioral phenotypes. Strikingly, all four mutants displayed a change in daytime locomotor activity suggesting that the ventral habenula may have an important role in modulating baseline activity. This is consistent with the role of ventral habenula in modulating levels of serotonin (Gabriel et al., 2009). All of these mutants, with the exception of *ntsr1*, also displayed reduction or increase in baseline neuronal activity within the habenula, within the ventral-most region of the structure.

The increase in arousal threshold in *aoc1* mutants together with the increase in locomotor activity was particularly intriguing in that it may represent a unique way in which habenula might exert its control on sleep. Aoc1 encodes an enzyme that deaminates histamine and might be involved in the regulation of histamine levels which is a known to promote wakefulness (Thakkar, 2011). In the absence of the enzyme that degrades extracellular histamine, the phenotypic effects observed in the *aoc1*-/- animals may be the result of an accumulation of extracellular histamine in the brain. These effects may be local within the neural circuitries of the habenula or global throughout the brain via the cerebrospinal fluid. Previous studies have observed reciprocal connections from the histaminergic centers of the fish brain to the habenula and vice versa (Kaslin and Panula, 2001; Stephenson-Jones et al., 2012). Further experiments investigating the global and local relative levels of histamine in mutants and wild type are required to address this hypothesis. Experimentations characterizing the detailed projections patterns of these *aoc1* positive neurons may reveal whether these neurons send or receive projections to and from histaminergic centers of the fish brain. It is also interesting to note that the selective expression of *aoc1* in the ventral habenula is a feature unique to non-mammalian brains (Amo et al., 2010). *Aoc1* is not expressed in the mammalian brain and consequently, the inactivating mechanism for histamine is dependent entirely on methylation through histamine N-methyltransferase.

In addition to baseline locomotor activity, further experimentation needs be performed to test the behavioral strategies of mutants when confronted with aversive environmental stimuli in the form of electric shocks, noxious heat or high salt conditions. Given the role of the ventral habenular cell types in behavioral responses to aversive environmental conditions (Andalman et al., 2019), it is particularly interesting to investigate these behaviors in the mutants of these marker genes. This represents an excellent opportunity to ask whether there is harmony between markers found by single cell analysis and the functional properties of that cell type within a circuit in the central nervous system. Furthermore, one of these genes *gpr139* has also been implicated in depression. If these mutants do have different coping strategies for aversive environmental conditions, the modulation of the *gpr139* receptor may represent an opportunity for the treatment of disorders such as depression. An extensive characterization of both behavior and brain activity through other behavioral phenotyping and calcium imaging (see Appendix Figure A.10) will be required to address these lines of inquiry.

3.6 MATERIALS AND METHODS

Zebrafish husbandry

Zebrafish larvae were maintained on 14hr:10hr light:dark cycle at 28°C. All protocols and procedures involving zebrafish were approved by the Harvard University/Faculty of Arts and Sciences Standing Committee on the Use of Animals in Research and Training (IACUC; Protocol #25-08). Larvae used for behavioral and brain activity screening were grown in 150mm Petri dishes in fish water containing methylene blue, at a density of less than 160 larvae per dish and debri was removed prior to 4dpf.

CRISPR-Cas9-mediated mutagenesis of *nts, ntsr1, aoc1* and *gpr139*

Mutants were generated by simultaneous microinjection of one to four guide RNAs(gRNAs), and approximately 0.5nL of 50µM Cas9 Protein into either EK/nacre (mitfa-/-) or gCaMP6s-/- embryos. Sequences of gRNAs used for each gene are listed in Table A.4. The resulting mosaic adults with germline mutations were then outcrossed to either EK or nacre wild type fish. The resulting offsprings (F1) were then subsequently mated to each other to generate homozygous mutants. All larvae were genotyped to generate stable lines for growing heterozygous and homozygous adults as well as for phenotypic characterizations. The primers used for genotyping are listed in Table A.4.

Generation of ISH probes

Flourescent RNA in situ hybridization was performed as described previously (Pandey et al., 2018) using 8dpf mitfa-/- larvae. To generate probes, gene fragments were amplified with Phusion polymerase (New England Biolabs, M0530L) using the primers listed in Table A.3. The Polymerase Chain Reaction (PCR)-amplified fragments were then cloned into pSC-A plasmid using Strataclone PCR Cloning Kit (Agilent, 240205), and used to transform the Strataclone competent cells. The transformed cells were plated overnight on Luria-Bertani (LB) agar plates. Colonies were selected by colony-PCR, cultured, mini-prepped and sent for sequencing. The resulting plasmids were then restricted with the appropriate restriction enzyme (Table S3), and purified using PCRclean up kit (Omega Cycle Pure Kit). The linearized vector was then used as a template to synthesize digoxigenin- or fluorescein labeled RNA probes using the RNA labeling kit (Roche). The transcription reactions were purified using Total RNA clean up kit (Omega, R6834), and the resulting RNA was quantified using Nanodrop and assessed on an agarose gel. The final product was then normalized to a concentration 50ng/mL in HM+ buffer (50% formamide, 5X Saline Sodium Citrate (SSC) buffer, 5 mgmL^{-1} torula RNA, 50 mgmL^{-1} heparin, 0.1% Tween 20) and stored at -20C until further use.

Behavioral Analysis

All assays were conducted with 48 to 96 larvae in a multi-well plate. Fish were transferred to 96 well 0.7ml plates (E&K Scientific) on the afternoon of 4dpf. Plates were filled with fish water containing methylene blue and transferred to ice until locomotion ceased. The plates were then gently sealed with optical adhesive films (Thermo Fisher Scientific) without the introducing any bubbles in wells. All experiments were conducted at 28°C unless otherwise noted. Locations of larvae were monitored with 1088x 660-pixed resolution using a Grasshopper3 camera (FLIR). Baseline locomotor activity were measured at 30 frames per second(fps) and stimulus responses were measured using one second long movies at 285 fps. Tap stimuli were delivered by computer-controlled increases in voltages to a solenoid attached to the apparatus. A series of tap strengths were used for each experiment. Each behavioral measurement were measured at least three times with three separate clutches of fish. Baseline locomotion measurements were performed separately from the experiments in which tap or dark flash stimuli were presented. Downstream behavioral analysis were performed as described before (Thyme et al., 2019).

Calcium imaging in the habenula (See Appendix Figure A.10)

Two-photon imaging of 6-7 dpf larvae was performed using a custom-built microscope in the Engert laboratory. Larvae were crossed, screened and selected for expression of GCaMP in the brain, and then embedded in 1.8% agarose. Under the microscope, fish were in complete darkness but were stimulated with electric shocks at 3V/cm at randomized time intervals. Calcium dynamics were measured across 8 slices separated by 7 microns to span across the dorso-ventral axis of the habenula. 500 frames were collected per slice at the speed of 1.4 seconds per frame. Each fish was imaged for a total of 3 hours.

In order to segment the imaging data into units that approximate cells, or coherent neuropil signals, we performed correlation-based segmentation as described previously (Portugues et al. 2014). For every unit that was generated, we calculated a metric to approximate its spontaneous activity. We defined this metric as the percentage of time that unit's calcium signal (defined as $\Delta F/F$) spent above a threshold value (Figure A, B). This threshold was uniquely calculated for each unit and defined as being three times the standard deviation of the absolute change in $\Delta F/F$ from one frame to the next above the mean absolute change in $\Delta F/F$ from one frame to the next.

After determining the activity of each detected region of interest from each fish, we sought to generate maps that represented the average activity across all fish from both of the two groups. To this end, we performed non-affine registration of stacks of average intensity over time from each fish onto the Z-Brain using the Computational Morphometry Toolkit (CMTK) software, similar to the registration done to generate the pERK maps. The transformation matrices generated by this registration were then used to map the above calculated spontaneous activity metrics onto their corresponding location of the Z-Brain. With all fish mapped to the same reference frame, we were then able to average the activity across fish.

Brain activity analysis

Images were registered to a standard zebrafish brain using CMTK. All phosphorylated-ERK images were normalized with a total- ERK stain. For each voxel, the Mann-Whitney U statistic Z score was calculated, comparing the mutant and control groups using MapMAPPING (Randlett et al., 2015). The significance threshold was set based on a false discovery rate where 0.05% of control pixels was called to be significant. Each experiment contained at least 20 mutant fish and 20 wild type fish. Any clutch smaller than 60 fish were discarded. All image processing analysis were parallelized in the Harvard Odyssey Cluster.

REFERENCES

Aizawa, H., Cui, W., Tanaka, K., and Okamoto, H. (2013a). Hyperactivation of the habenula as a link between depression and sleep disturbance. Front Hum Neurosci *7*, 826.

Aizawa, H., Yanagihara, S., Kobayashi, M., Niisato, K., Takekawa, T., Harukuni, R., McHugh, T.J., Fukai, T., Isomura, Y., and Okamoto, H. (2013b). The synchronous activity of lateral habenular neurons is essential for regulating hippocampal theta oscillation. J Neurosci *33*, 8909-8921.

Almeida, A.P., and Beaven, M.A. (1981). Phylogeny of histamine in vertebrate brain. Brain Res *208*, 244-250.

Amo, R., Aizawa, H., Takahoko, M., Kobayashi, M., Takahashi, R., Aoki, T., and Okamoto, H. (2010). Identification of the zebrafish ventral habenula as a homolog of the mammalian lateral habenula. J Neurosci *30*, 1566-1574.

Andalman, A.S., Burns, V.M., Lovett-Barron, M., Broxton, M., Poole, B., Yang, S.J., Grosenick, L., Lerner, T.N., Chen, R., Benster, T., *et al.* (2019). Neuronal Dynamics Regulating Brain and Behavioral State Transitions. Cell *177*, 970-985 e920.

Burgess, H.A., and Granato, M. (2007). Modulation of locomotor activity in larval zebrafish during light adaptation. J Exp Biol *210*, 2526-2539.

Chen, A., Singh, C., Oikonomou, G., and Prober, D.A. (2017). Genetic Analysis of Histamine Signaling in Larval Zebrafish Sleep. eNeuro *4*.

Dunn, E.C., Wiste, A., Radmanesh, F., Almli, L.M., Gogarten, S.M., Sofer, T., Faul, J.D., Kardia, S.L., Smith, J.A., Weir, D.R., *et al.* (2016). Genome-Wide Association Study (Gwas) and Genome-Wide by Environment Interaction Study (Gweis) of Depressive Symptoms in African American and Hispanic/Latina Women. Depress Anxiety *33*, 265-280.

Gabriel, J.P., Mahmood, R., Kyriakatos, A., Soll, I., Hauptmann, G., Calabrese, R.L., and El Manira, A. (2009). Serotonergic modulation of locomotion in zebrafish: endogenous release and synaptic mechanisms. J Neurosci *29*, 10387-10395.

Jefferis, G.S., Potter, C.J., Chan, A.M., Marin, E.C., Rohlfing, T., Maurer, C.R., Jr., and Luo, L. (2007). Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. Cell *128*, 1187-1203.

Kaslin, J., and Panula, P. (2001). Comparative anatomy of the histaminergic and other aminergic systems in zebrafish (Danio rerio). J Comp Neurol *440*, 342-377.

Kokel, D., Bryan, J., Laggner, C., White, R., Cheung, C.Y., Mateus, R., Healey, D., Kim, S., Werdich, A.A., Haggarty, S.J., *et al.* (2010). Rapid behavior-based identification of neuroactive small molecules in the zebrafish. Nat Chem Biol *6*, 231-237.

Kononoff, J., Kallupi, M., Kimbrough, A., Conlisk, D., de Guglielmo, G., and George, O. (2018). Systemic and Intra-Habenular Activation of the Orphan G Protein-Coupled Receptor GPR139 Decreases Compulsive-Like Alcohol Drinking and Hyperalgesia in Alcohol-Dependent Rats. eNeuro *5*.

Liu, C., Bonaventure, P., Lee, G., Nepomuceno, D., Kuei, C., Wu, J., Li, Q., Joseph, V., Sutton, S.W., Eckert, W., *et al.* (2015). GPR139, an Orphan Receptor Highly Enriched in the Habenula and Septum, Is Activated by the Essential Amino Acids L-Tryptophan and L-Phenylalanine. Mol Pharmacol *88*, 911-925.

Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C.*, et al.* (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science *362*.

Nohr, A.C., Shehata, M.A., Hauser, A.S., Isberg, V., Mokrosinski, J., Andersen, K.B., Farooqi, I.S., Pedersen, D.S., Gloriam, D.E., and Brauner-Osborne, H. (2017). The orphan G protein-coupled receptor GPR139 is activated by the peptides: Adrenocorticotropic hormone (ACTH), alpha-, and beta-melanocyte stimulating hormone (alpha-MSH, and beta-MSH), and the conserved core motif HFRW. Neurochem Int *102*, 105-113.

Pandey, S., Shekhar, K., Regev, A., and Schier, A.F. (2018). Comprehensive Identification and Spatial Mapping of Habenular Neuronal Types Using Single-Cell RNA-Seq. Curr Biol *28*, 1052-1065 e1057.

Randlett, O., Wee, C.L., Naumann, E.A., Nnaemeka, O., Schoppik, D., Fitzgerald, J.E., Portugues, R., Lacoste, A.M., Riegler, C., Engert, F.*, et al.* (2015). Whole-brain activity mapping onto a zebrafish brain atlas. Nat Methods *12*, 1039-1046.

Rihel, J., Prober, D.A., Arvanites, A., Lam, K., Zimmerman, S., Jang, S., Haggarty, S.J., Kokel, D., Rubin, L.L., Peterson, R.T.*, et al.* (2010). Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. Science *327*, 348-351.

Rohlfing, T., and Maurer, C.R., Jr. (2003). Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. IEEE Trans Inf Technol Biomed *7*, 16-25.

Schwelberger (2004a). Histamine N-methyltransferase (HNMT) enzyme and gene. Histamine: Biology and medical aspects, 53-59.

Schwelberger (2004b). Diamine oxidase(DAO) enzyme and gene. . Histamine: Biology and medical aspects.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron *92*, 342-357.

Stephenson-Jones, M., Floros, O., Robertson, B., and Grillner, S. (2012). Evolutionary conservation of the habenular nuclei and their circuitry controlling the dopamine and 5-hydroxytryptophan (5-HT) systems. Proc Natl Acad Sci U S A *109*, E164-173.

Sundvik, M., Kudo, H., Toivonen, P., Rozov, S., Chen, Y.C., and Panula, P. (2011). The histaminergic system regulates wakefulness and orexin/hypocretin neuron development via histamine receptor H1 in zebrafish. FASEB J *25*, 4338-4347.

Sundvik, M., and Panula, P. (2012). Organization of the histaminergic system in adult zebrafish (Danio rerio) brain: neuron number, location, and cotransmitters. J Comp Neurol *520*, 3827-3845.

Thakkar, M.M. (2011). Histamine in the regulation of wakefulness. Sleep Med Rev *15*, 65-74.

Thyme, S.B., Pieper, L.M., Li, E.H., Pandey, S., Wang, Y., Morris, N.S., Sha, C., Choi, J.W., Herrera, K.J., Soucy, E.R.*, et al.* (2019). Phenotypic Landscape of Schizophrenia-Associated Genes Defines Candidates and Their Shared Functions. Cell *177*, 478-491 e420.

Wang, D., Li, Y., Feng, Q., Guo, Q., Zhou, J., and Luo, M. (2017). Learning shapes the aversion and reward responses of lateral habenula neurons. Elife *6*.

Chapter 4: Cell Type Landscape of the Forebrain and Cell Type-specific Dissection of Forebrain Mutants

---

4.1 PREFACE

A part of the chapter on the *znf536* -/- fish was published in Cell on April 4, 2019. The rest of the chapter is being prepared for publication. I designed, performed and interpreted the experiments with input from Alexander F. Schier. Summer Thyme generated the CRISPR mutants. Alexander F. Schier supported the project.

4.2 ABSTRACT

The zebrafish telencephalon is composed of a set of highly specialized subregions that are known to regulate learning, memory, social behavior and emotion. However, its cell types remain poorly described. We used single cell RNAseq and in situ hybridization to describe a comprehensive repertoire of cell types in the zebrafish telencephalon. Single cell transcriptomes of ~35,000 cells from larval and adult telencephalon were used to delineate 15 neuronal types in the larval telencephalon and 18 neuronal types in the adult telencephalon. Some larval neuronal types were not present in the adult telencephalon, some others were largely similar to their adult counterparts and yet others diverge into multiple types in the adult telencephalon. Lastly, single cell analysis of znf536-/- telencephalons revealed a loss of a specific cell group of *uts1* neurons in the subpallium. Our work provides a comprehensive transcriptional analysis of the cell

types in the larval and adult telencephalon and forms a resource for dissecting the development and function of the zebrafish telencephalon.

4.3 RESULTS

Graph Clustering Identifies Cell Types in the Larval Zebrafish Telencephalon

The telencephalon encompasses the frontal lobe of the zebrafish brain and has important functions in learning, memory, social behavior, and decision-making (Aoki et al., 2013; Cheng et al., 2014; Lal et al., 2018; Stednitz et al., 2018). To analyze the molecular features of the cells from the zebrafish forebrain, we obtained 12, 485 single cells from larval telencephalon using 10X. Standard computational pipelines, quality and tissue filtering were then used to align the raw sequencing data to the zebrafish transcriptome and derive a gene expression matrix of 19,469 genes across 11,855 cells. For the larval telencephalon, 2000 highly variable genes across the datasets were selected using a UMI based approach that ranks genes based on a deviation from a null statistical model that described the technical variation in each gene given its expression level (Pandey et al., 2018). These genes were then used to identify 50 significant principal components PCs that were used to partition the cells into 27 transcriptionally distinct clusters using smart local moving community detection algorithm (Waltman, 2013) using Seurat R package (Satija et al., 2015). Each cluster was then subjected to iterative clustering to assess additional heterogeneity, which was visualized in two dimensions using t-distributed stochastic neighborhood embedding(t-SNE) (Figure 4.1A). Each of these clusters were then evaluated for differential gene expression analysis to identify cluster-specific markers for the two main parts of the telencephalon, pallium and subpallium respectively (Figure 4.1C-D).

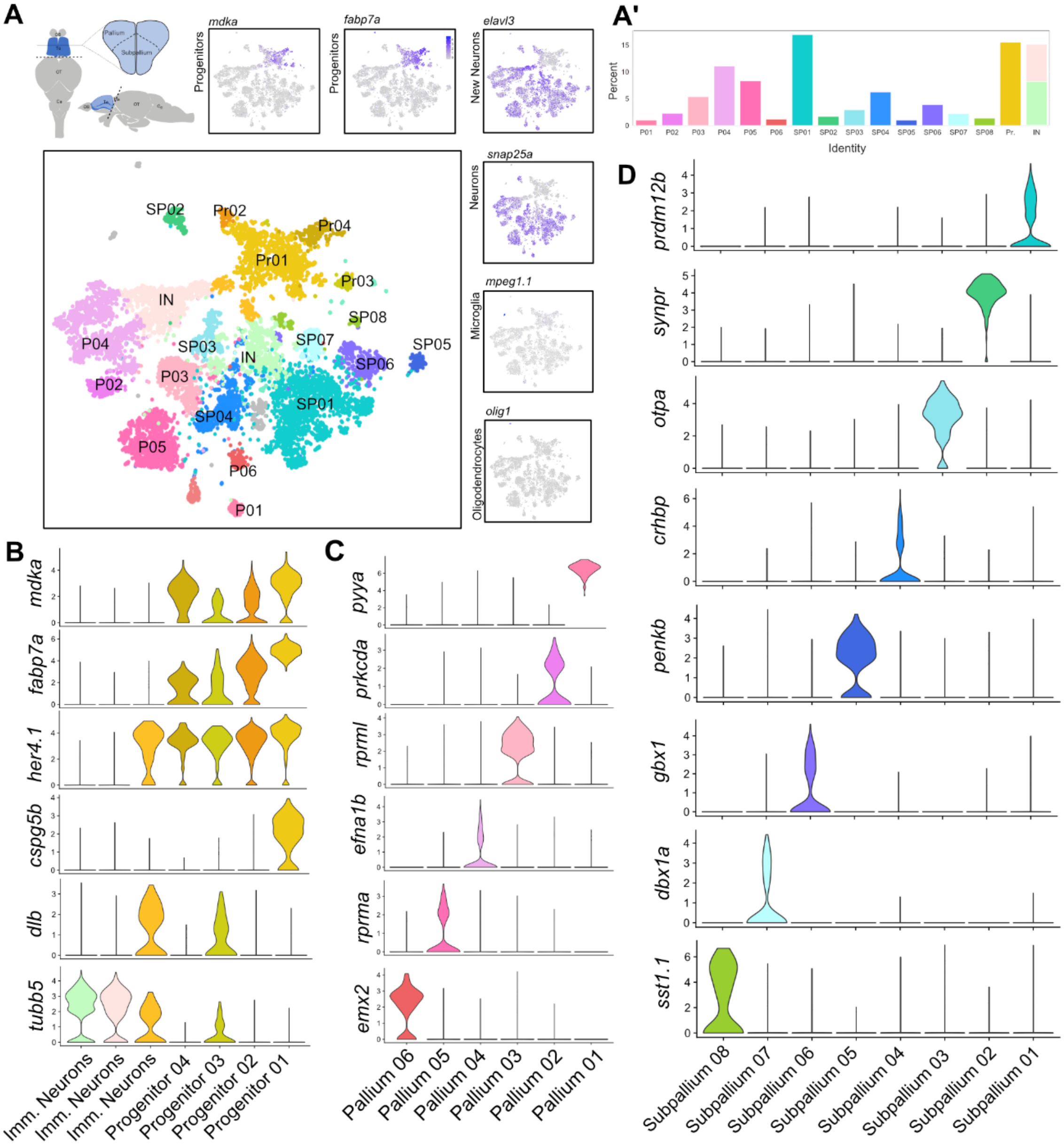


Figure 4.1. Unbiased Clustering of Telencephalic Neurons Identifies 6 Pallium and 8 Subpallium Neuronal Types.

(A) 2D visualization of single cell clusters in the larval telencephalon. Cluster Labels are as follows: P01, P02, etc: Pallium; SP01, SP02, etc: Subpallium; PoA: Preoptic Area; OB: Olfactory Bulb; Pr: Progenitors; IN: Immature Neurons. Top left: Schematic of the zebrafish forebrain showing the anatomical subdivisions

Figure 4.1 (Continued). corresponding to the pallium and subpallium. Top and side panels: Expression patterns of neuronal and non-neuronal genes among the telencephalic clusters. A': Bar chart displaying the percentage of total dataset represented in every cluster, showing the abundance of each cell type found by clustering analysis.

(B) Gene expression profiles of select cluster-specific markers for progenitors and immature neurons identified through the differential gene expression analysis.

(C) Gene expression profiles of select cluster-specific markers for pallium neurons identified through the differential gene expression analysis.

(D) Gene expression profiles of select cluster-specific markers for subpallium neurons identified through the differential gene expression analysis.

Of the 27 clusters we identified in our analysis of larval forebrain, 18 clusters were neuronal and 9 clusters were non-neuronal (Figure 4.1). Among the non-neuronal cells, we found clusters expressing prototypical markers for microglial cells, oligodendrocytes and ependymal cells (Figure 4.1A, top and side panels). We also found progenitors, including radial glial cells that express markers such as *fabp7a, gfap* and proneural progenitors that express markers such as *neurog1* (Figure 4.1B). Among the clusters we label as radial glial cells, a single cluster also expresses canonical astrocyte markers such as *slc1a3b* and *slc1a2b*. In addition, they also express *mfge8a*, a marker recently reported to be specific for telencephalic astrocytes (Zeisel et al., 2018)

Among the *tubb5*+ immature neurons, we found two distinct types in the larval telencephalon: one expressing broad transcriptions factors essential for pallial development such as *bhlhe22* and *eomesa* and another expressing broad sub-pallial transcription factors such as *dlx2a and dlx5a* (Figure A.8). To assign rough regional identity to the mature neuronal clusters, we identified a host of markers for each cluster. A large majority of clusters were defined by single genes. We assigned clusters as belonging to pallium and subpallium based on the expression patterns of these gene markers as defined by *in situ* hybridization and the expression of known pallium and sub-pallium markers among our clusters (Figure 4.2, Figure A.8). Among the neuronal clusters we also found one cluster of neurons belonging to the olfactory bulb and one belonging to the preoptic area, which will be removed from further discussion.

We found 7 clusters belonging to the pallium, the dorsal region of the telencephalon that is thought to contain structures homologous to the mammalian cortex, hippocampus and some sections of the amygdala (Ganz et al., 2014). In correspondence with this, we found genes enriched in the cortex and hippocampus such as *bcl11ba, zbtb18, fezf1* to be enriched in different clusters belonging to the pallium (Figure A.8A). Some other classical cortical markers such as *satb2* were absent in the zebrafish telencephalon. Consistent with this and in contrast to human studies, our others' previous work shows that *satb2-/-* zebrafish have no behavioral phenotypes (Thyme et al., 2019; Zarate and Fish, 2017).

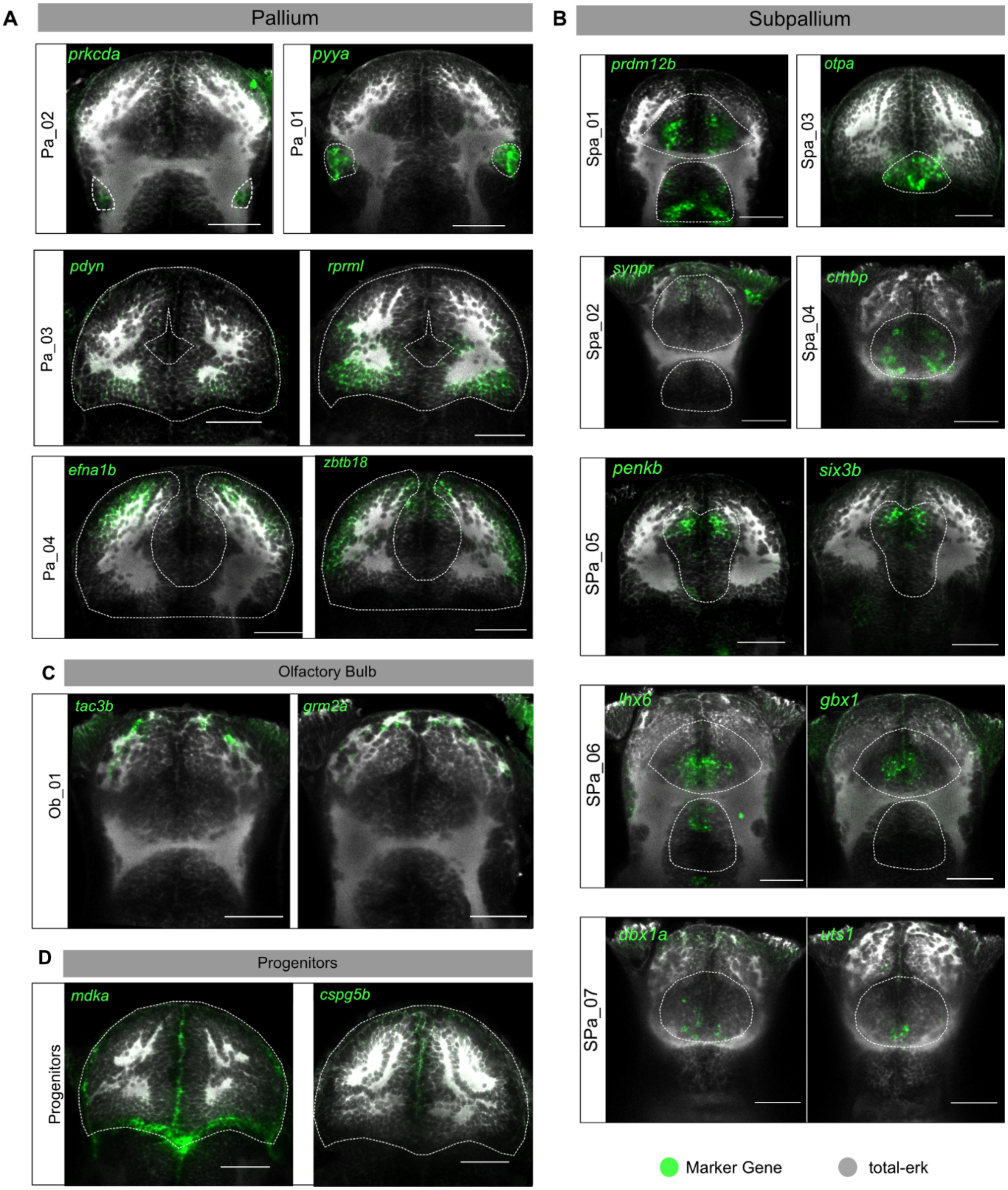


Figure 4.2. Spatial Distribution of the Pallium and Subpallium Neuronal Types in the Larval Telencephalon.

(A) *In vivo* expression patterns of cluster-specific marker genes for selected sub-clusters in the pallium. RNA-FISH (green) was performed with a total-Erk (pale gray) co-stain for anatomy.

(B) *In vivo* expression patterns of cluster specific marker genes for selected clusters belonging to the subpallium. RNA-FISH (green) was performed with a total-Erk (pale gray) co-stain for anatomy.

Figure 4.2 (Continued). (C) *In vivo* expression patterns of cluster specific marker genes for the cluster belonging to the olfactory bulb. RNA-FISH (green) was performed with a total-Erk (pale gray) co-stain for anatomy.

(D) *In vivo* expression patterns of cluster specific marker genes for two progenitor clusters. RNA-FISH (green) was performed with a total-Erk (pale gray) co-stain for anatomy.

We identified 8 clusters belonging to the sub-pallium, the ventral region of the telencephalon which is thought to contain structures homologous to the mammalian striatum and basal ganglia. In alignment with this, we found a number of striatum marker genes such as *isl1b, drd2b, penkb, synpr* expressed specifically within selected clusters within the subpallium (Figure A.8B). Also notable among the sub-pallium clusters were *otpa*+ SP03 that is located in the dorsal-most and posterior region of the sub-pallium and is thought to be a region homologous to the medial amygdala (Biechl et al., 2017). We also captured functionally distinct classes of neurons among our single cell clusters. For instance, SP06 neurons positive for lhx8a+ were recently found to be required for the control of behaviorally-driven social orienting in zebrafish (Figure 4.1D) (Stednitz et al., 2018). Taken together, these results provide a catalog of markers for neuronal types belonging to the pallium and sub-pallium.

Spatial localization and validation of neuronal types in the zebrafish telencephalon

To localize the single cell clusters obtained from the larval forebrain, we examined the expression patterns of the assigned marker genes using in situ hybridization (Figure 4.2). Similar to previously published work (Pandey et al., 2018), we also found that the neuronal types in the pallium and sub-pallium were largely regionalized. In correspondence with the cluster abundances in the single cell data, Pa04 occupied the largest expression domains within the pallium towards the anterior half of the dorsal telencephalon (Figure 4.2A, bottom panel). On the other hand, Pa03 characterized by the expression of *pdyn* and *rprml* were localized to the posterior half of the dorsal telencephalon (Figure 4.2A, middle panel). Two of the smaller clusters within the pallium Pa01 and Pa02 were localized more laterally, spread along the dorsoventral axis (Figure 4.2A, top panel).

Subpallium neuronal types were also regionalized along the anterior-posterior and dorso-ventral axis in the ventral part of the telencephalon. Neuronal clusters expressing striatum like markers such as SP05 (*penkb*+), SP02(*synpr*+) were located in the anterior half of the subpallium along the dorso-ventral axis (Figure 4.2B). SP04, SP06 and SP07 are located in the medial part along the anterior posterior axis along different levels of the dorso-ventral axis. Lastly otpa+ SP03 was located in the posterior half of the subpallium along the dorsal most part of the subpallium.

Together, these results provide validation and spatial localization for the majority of neuronal clusters found in the telencephalon through our single cell analysis.

Comparison of Larval Telencephalon to Adult Telencephalon

The telencephalon undergoes a massive growth, neurogenesis and functional maturation from developing larvae to mature adults. Indeed, several behaviors mediated by neuronal ensembles in the telencephalon develop later in development including social behaviors and fear learning (Dreosti et al., 2015; Lal et al., 2018). To assess the conservation and retention of neuronal types between the larvae and adult telencephalon, we also captured and sequenced cells from the adult telencephalon in two batches of single cell experiments (Figure A.9B). Post quality filtering and batch correction, we obtained 27,125 single-cell profiles at a median depth of 2488 UMI and 1070 genes per cell (Figure A.9A). Using the same clustering technique as before, we recovered 32 clusters and enriched cluster-specific markers (Figure 4.3A). Out of the 32 clusters, 2 were obtained from a single experimental batch and belonged to the habenula. These were removed from further analysis.

Among the rest of the cells in the dataset, we found both neuronal and non-neuronal cells. (Figure 4.3A, side and top panels). The neuronal clusters were labeled post-hoc after comparison to the larval clusters (see below). In addition to assigning specific markers to adult clusters, we also divided the neuronal cells in the adult data into excitatory and inhibitory cells by sub-setting cells that are high in *gad2* and *vglut2* respectively and clustering them into distinct types. Among the inhibitory neurons, we found 5 distinct groups expressing *penkb, pyya, lhx8a, adcyap1b, sst1.1* and *galn* respectively. A smaller proportion of cells expressed *slc17a6a* and *slc17a6b* at a high level. Hence, we only recovered 3 sub-clusters among the *slc17a6a*/*slc17a6b*+ excitatory neurons although there are probably more *in vivo* (Figure 4.3C). Interestingly, one of the excitatory neuron clusters was *pvalb7*+, which are typically thought to be GABAergic neurons.

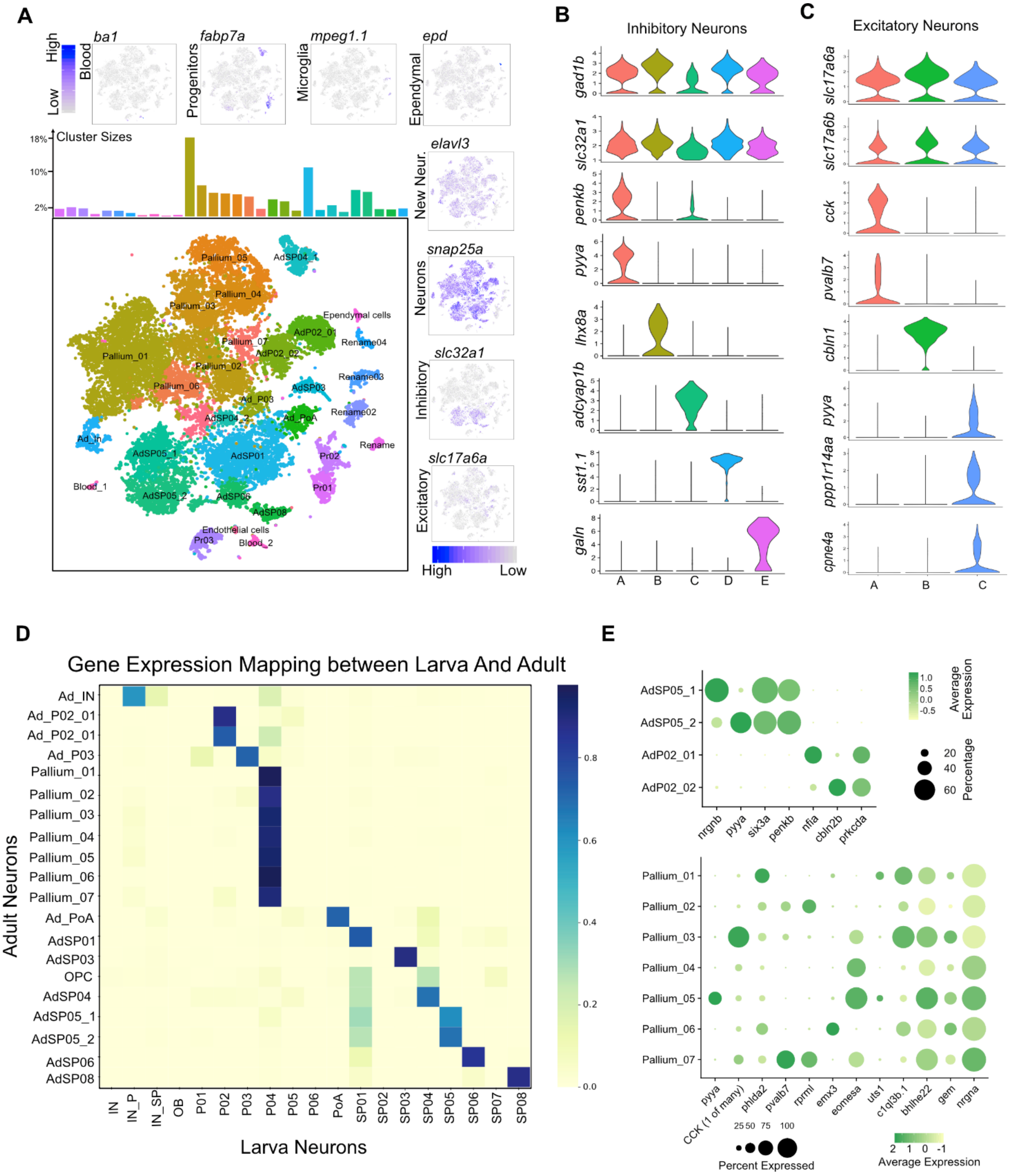


Figure 4.3. Comparison of Larval Cell Types with Adult Cell Types.

(A) 2D visualization of the single cell clusters in the adult forebrain. Top and side panels: Expression patterns of neuronal and non-neuronal genes among the telencephalic clusters. Bar chart displays the percentage of total dataset represented in every cluster, showing the abundance of each cell type found by clustering analysis.

Figure 4.3 (Continued). (B) Gene expression profiles for clusters of inhibitory neurons in the telencephalon

(C) Gene expression profiles for clusters of excitatory neurons in the telencephalon.

(D) Heatmap showing the proportion of adult cells that were classified to larval cluster labels.

(E) Gene expression profiles of new and old marker genes for adult clusters that map to multiple neuronal types in the larval telencephalon.

To systematically compare the neuronal clusters between the larva and adult telencephalon, we trained a gradient boosted tree-based model to map gene expression signatures between two datasets (Figure A.9C). In contrast to our previous study, we observed less correspondence between larval and adult clusters (Pandey et al., 2018). We observed that 8 out of 20 neuronal clusters in the larval telencephalon mapped 1:1 between larva and adult (Figure 4.3D). Most of these clusters were in the subpallium region of the telencephalon, suggesting that the subpallium generates distinct neuronal types with stable identities earlier in development than the pallium. *Penkb*+ SP05 diverges into two adult types that are both positive for *penkb* and *six3a* but are differentiated by the expression of *nrgnb* and *pyya* respectively. SP02 (*synpr*+) and SP07 (*dbx1a*+) are lost in the adult telencephalon.

Among the larval pallium neuronal types, only *rprml*+ P03 mapped 1:1 to a single adult cluster. *Prkcda*+ P02 diverged into two clusters that were distinguished by the expression of *nfia* and *cbln2b* (Figure 4.3E). The other larval clusters from the pallium were largely absent in the adult pallium. All of the rest of the adult pallium neuronal types mapped to a single larval cluster P04. All of these clusters express the canonical markers for larval P04 including *nrgna, gem* and *bhlhe22* but are distinguished by a combinatorial expression patterns of a number of genes (Figure 4.3E). Our results demonstrate that the telencephalon, particularly the pallium, grows not just in size during development but also establishes new cell type domains from larva to adult. Of the new domains specific to the adult telencephalon, we validated the specific expression of parvalbumin positive cluster of neurons that was shown to be important for the retrieval of an aversive reinforcement learning behavior (Figure 4.4A)

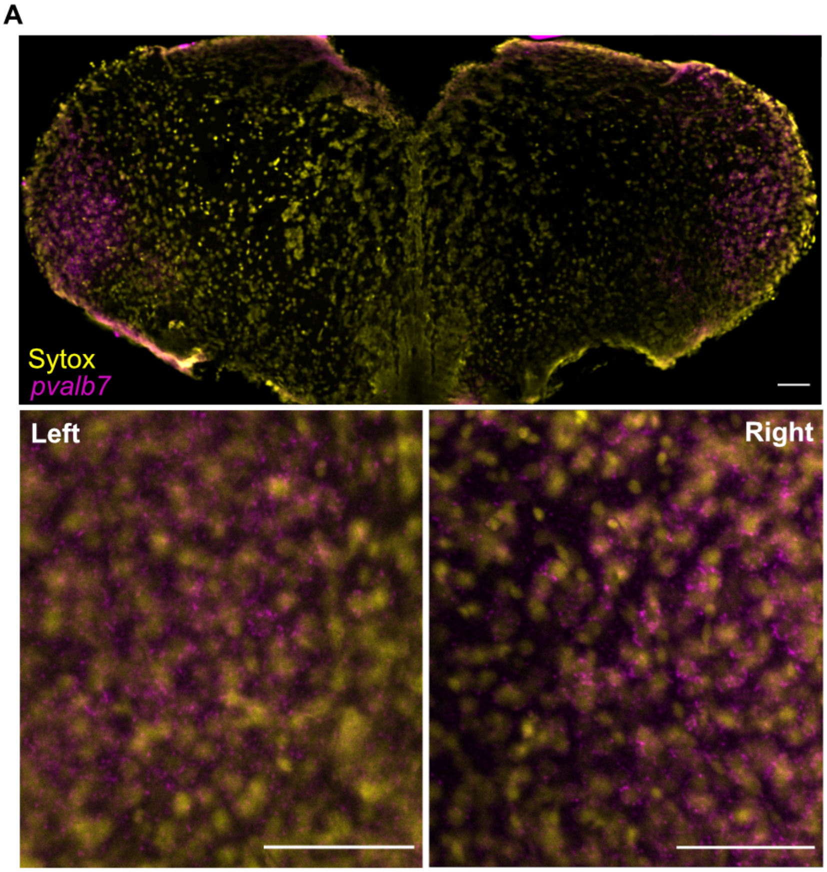


Figure 4.4: Expression Pattern of Parvalbumin in the Adult Telencephalon.

(A) In situ hybridization for *pvalb7* in the adult telencephalon. Left and right insets are zoomed in regions from the left and right half of the telencephalon. *Pvalb7* expression not detected in the larval telencephalon (data not shown).

Comparison of cell type landscape of wild type and znf536-/- fish

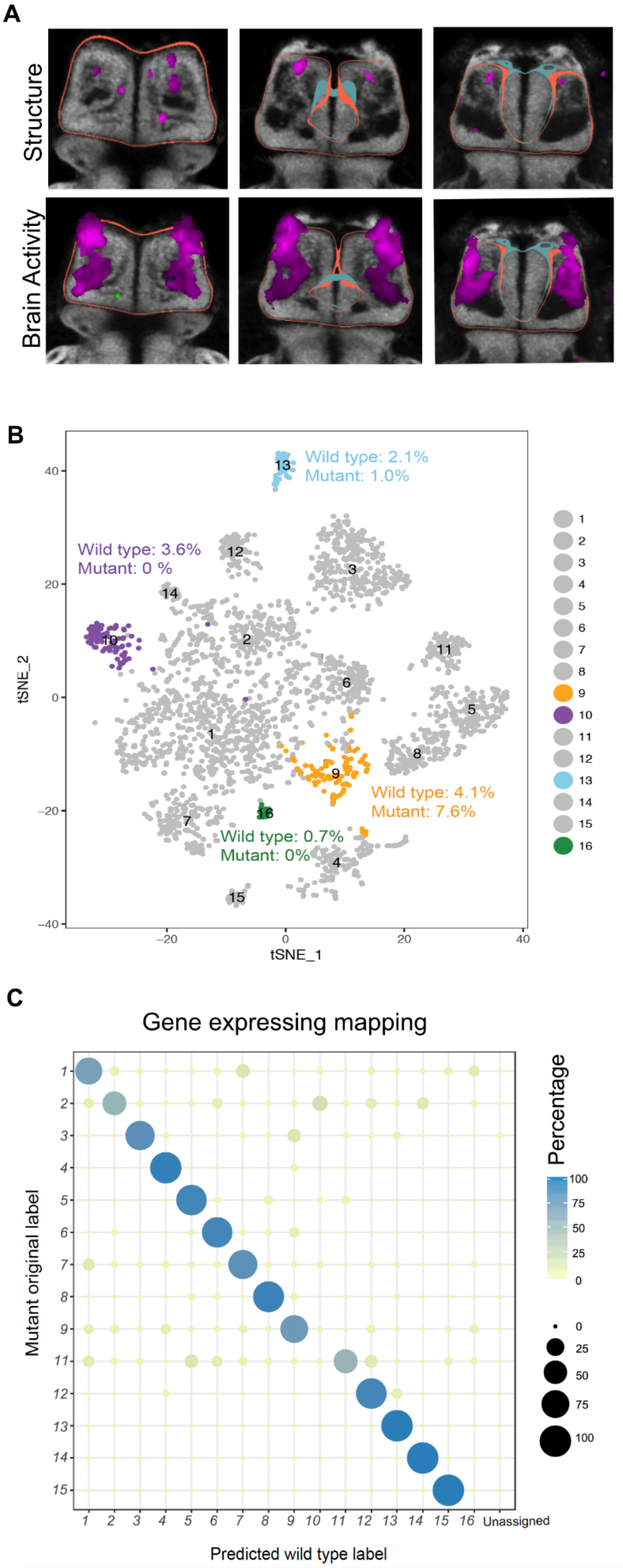
The transcription factor *znf536* is expressed in the forebrain but is largely uncharacterized functionally. It is also associated with multiple neuropsychiatric disorders (Schizophrenia Working Group of the Psychiatric Genomics, 2014; Winham et al., 2014). Zebrafish mutants for this gene exhibited volume loss and activity differences in the pallium region of the forebrain (Figure 4.5A). To uncover what neuron types may be missing in forebrain, we profiled the single cell transcriptomes of dissected mutant and wild-type sibling forebrains (Figure 4.5). Using a random forest analysis, we compared the gene expression signatures and cluster compositions between from wild type and *znf536-/-* animals (Figure 4.5C). This comparison revealed a loss (*uts1*+ subpallium cluster and *penkb*+ subpallium cluster) and reduction of (*tac3b*+ pallium cluster) cell types (Figure 4.5B, Figure 4.5C). All of markers of these clusters are neuropeptides involved in stress and social behaviors (Neufeld-Cohen et al., 2010) Corresponding with the loss of the *uts1* population is an increase in a pool of immature neurons (*tubb5*+*),* suggesting a potential developmental mechanism for the phenotype (Figure 4.5C). The *tac3b* population is additionally marked by *grm2a* (Figure S6), which is a potential schizophrenia therapeutic target (Griebel et al., 2016; Li et al., 2015; Mossner et al., 2008). The identification of these specifically affected types of neurons illustrates that single-cell RNA sequencing allows the dissection of the biological processes regulated by schizophrenia-associated genes.

Figure 4.5. Cell type specific phenotypic dissection of *znf536.*

(A) Brain activity and structural differences in the znf536-/- forebrains.

(B) t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of wild-type single-cell clusters obtained by clustering of 6-dpf forebrain cells. Clusters with substantial differences in *znf536* mutants are highlighted in orange, purple, and blue. Cluster counts in mutant and wild type are expressed as percent of the total cell number for each sample.

(C) Dotplot (confusion matrix) showing the proportion of cells in the *znf536* mutant forebrain that were classified to wild-type cluster labels. Each mutant forebrain type was assigned to a wild-type cluster label if > 13% of the trees in the Random Forest (RF) model contributed to majority vote.


A
Structure
Brain Activity
B
Wild type: 2.1%
Mutant: 1.0%
Wild type: 3.6%
Mutant: 0 %
Wild type: 4.1%
Mutant: 7.6%
Wild type: 0.7%
Mutant: 0%
tSNE_1
tSNE_2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
C
Gene expressing mapping
Mutant original label
Predicted wild type label
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 Unassigned
Percentage
0 25 50 75 100

4.4 DISCUSSION

Using a comprehensive single cell analysis in both larva and adult, we found 18 neuronal types in larva and 20 neuronal types in the adult telencephalon. For each of these neuronal types, we found dozens of novel molecular markers. We also used RNA-FISH and image registration to spatially localize neuronal types, a key step in linking molecular profiles to physiological and behavioral features. Next, we found remarkable divergence between the cell types in larval and adult telencephalon. Our results demonstrated that while there is a conservation of neuronal types between the larval and adult subpallium, a large portion of the pallium establishes new cell type domains going from larval to adult animals. Of the new domains specific to the adult telencephalon, an important one is a parvalbumin positive cluster of neurons that was previously known to be important for the retrieval of an aversive reinforcement learning behavior (Figure 4.3E, Figure 4.4A). Other adult-specific domains may also have important implications for the emergence of social behaviors and learning later in development.

Next, we analyzed the telencephalons of *znf536*-/- animals in which we found large loss of structure in the telencephalon. Using a comparative analysis of cell types between the wild type and mutant telencephalon, we discovered a loss of specific population of peptidergic neurons, including *tac3b, penkb* and *uts1*. This was accompanied by an increase in the number of immature neurons in the znf536-/- telencephalons, suggesting that *znf536* might be a transcription factor important for the maturation of the immature neurons into these peptidergic subtypes. A careful time course RNA-seq analysis building with developmental trajectories might help address this hypothesis.

Overall, this chapter provides a comprehensive survey of cell types in the zebrafish telencephalon and provides a blueprint for the dissection of cell type specific molecular underpinnings of brain activity and behavioral phenotypes in mutant animals.

4.5 MATERIALS AND METHODS

Experimental Model and Subject Details.

Larvae and adult fish were maintained on 14 hours: 10 hours light: dark cycle at 28ºC. All protocols and procedures involving zebrafish were approved by the Harvard University/Faculty of Arts & Sciences Standing Committee on the Use of Animals in Research and Teaching (IACUC; Protocol #25-08). 6 days post fertilization (dpf) larval and ~1-year old adult zebrafish were used. Animals were anesthetized in 0.2% tricaine and rapidly euthanized by immersion in ice water for 5 minutes before dissection.

METHOD DETAILS

Cell Isolation and Single cell RNA-seq

6 dpf larval forebrains were dissected from six wild type fish in Neurobasal (Thermo Fisher Scientific 21103049) supplemented with 1X B27 (ThermoFisher Scientific

17504044) and promptly dissociated with the Papain Dissociation Kit (Worthington Biochemical Corporation LK003150) with the following modifications. Larval forebrains were incubated in 20 units/mL papain for 12 minutes at 37ºC. The cells were dissociated by gentle trituration 20 times and spun at 300xg for 5 minutes. The cells were resuspended in 1.1 mg/mL papain inhibitor in Earle's Balanced Salt Solution (EBSS), spun and 300xg for 5 minutes. The resulting pellet was then washed in Neurobasal supplemented with B27 before final resuspension in 50 µL PBS + 200 µg/mL BSA. To avoid loss of cells during each wash, the supernatant was saved and spun a second time to recover the remaining cells. The cells were then mixed with the final suspension to increase cell count. Viability and cell number was assessed by trypan blue staining on 10ul of the sample. If viability was greater than 80%, the cells were loaded on the 10X Chromium system at a concentration of approximately 150 cells/ µL. Libraries were prepared according the manufacturer's instructions. Single cell suspensions from adult forebrains were generated with the same protocol. However, only 1 forebrain was used per sample and was incubated in 20 units/m papain for 30 minutes at 37ºC.

Flourescent RNA In situ hybridization

Florescent RNA in situ hybridizations were performed exactly as previously described (Pandey et al., 2018; 15 Ronneberger et al., 2012) using 10 dpf *mitfa*$^{-/-}$ larvae. To generate probes, gene fragments were amplified from with Phusion polymerase (New England Biolabs, 18 M0530L) using the primers that are listed in Suppplemental Table 3. The Polymerase Chain Reaction (PCR)-amplified fragments were then cloned into pSC-

A plasmid using Strataclone PCR Cloning Kit (Agilent, 240205), and used to transform the Strataclone competent cells. The transformed cells were plated overnight on Luria-Bertani (LB) agar plates. Colonies were selected by colony-PCR, cultured, mini-prepped and sent for sequencing. The resulting plasmids were then restricted with the appropriate restriction enzyme (Supplementary Table 3), and purified using PCR-clean up kit (Omega Cycle Pure Kit). The linearized vector was then used as a template to synthesize digoxigenin- or fluorescein-labeled RNA probes using the RNA labeling kit (Roche). The transcription reactions were purified using Total RNA clean up kit (Omega, R6834), and the resulting RNA was quantified using Nanodrop and assessed on an agarose gel. The final product was then normalized to a concentration 50ng/μL in HM+ buffer (50% formamide, 5X Saline Sodium Citrate (SSC) buffer, 5 $mgmL^{-1}$ torula RNA, 50 μgmL^{-1} heparin, 0.1% Tween 20) and stored at -20ºC until further use.

Computational Methods for Data Analysis

Alignment, quantification, filtering and clustering analysis for single cell datasets

Raw sequencing data was converted to matrices of expression counts using the cellranger software provided by 10X genomics[1]. Briefly raw BCL files from the Illumina NextSeq or HiSeq were demultiplexed into paired-end, gzip-compressed FASTQ files for each channel using “cellranger mkfastq”. Both pairs of FASTQ files were then provided as input to “cellranger count” which partitioned the reads into their cell of origin

---

[1] https://support.10xgenomics.com/single-cell-gene-expression/software/ pipelines/latest /what-is-cell-ranger

based on the 14 base-pair (bp) cell barcode on the left read. Reads were aligned to a zebrafish reference transcriptome (ENSEMBL Zv10, release 82 reference transcriptome), and transcript counts quantified for each annotated gene within every cell. Here, the 10 base-pair unique molecular identifier (UMI) on the left read was used to collapse PCR duplicates, and accurately quantify the number of transcript molecules captured for each gene in every cell. Both cellranger mkfastq and cellranger count were run with default command line options. This resulted in an expression matrix (genes x cells) of UMI counts for each sample. Using the expression matrix, cells were filtered to remove those that contained less than 200 genes and those in which > 6% of transcript counts were derived mitochondrial-encoded genes. Similarly, genes detected in less than 5 cells were removed. Among the remaining cells, the median number of UMIs per cell was 2,279 and the median number of genes was 1,319 for larval data. The same for adult data was 1,614 UMI/cell and 709 genes/ cell, respectively (Figure A.7A, A.7B, A.9A and A.9B).

We used a linear regression model to correct for batch effects in the gene expression matrix using the RegressOut function in the Seurat R package, and used the residual expression values for further analysis. The residual matrix was then scaled, centered and used for the selection of variable genes, PCA and clustering.

All cells derived from regions other than the forebrain, such as habenula and olfactory bulb, were removed from further analysis. The residual matrix was then scaled, centered and used for further analysis. To select highly variable genes, we used a combination of a UMI based method described recently (Pandey et al., 2018) and Seurat's (Satija et al., 2015) variable gene selection approach. The resulting expression matrix across the highly variable genes was then used to perform dimensionality reduction and clustering using Seurat.

Comparison of cluster specific signatures between larval and adult dataset using gradient boosted trees

To evaluate the correspondence between clusters found in the larval telencephalon and adult telencephalon, we trained a multi-class classifier using the gradient boosted trees (GBT) on the larval dataset. Gradient boosted trees is an additive and stagewise tree-based machine learning algorithm that improves an ensemble of weak learners in a sequential manner to increase predictive power (J. Friedman, 2001).

We composed a training set for the classifier by taking a stratified sampling of cells from the 19 neuronal clusters from the larval dataset. 90% of the cells was used for training and 10% were used for testing the classifier. The classifier was built on the most variable genes across both the larval and adult dataset using the package sklearn in Python. Hyperparameters for the model were selected using grid search with 5-fold cross validation. The trained classifier was then used to assign a cluster label for remaining 10% of the data (Figure 4.1C). Cells in the test set were accurately mapped to their correct classes by the trained classifier at an average accuracy 93% and precision of 92.5%. Precision-recall curves were drawn separately for each class to evaluate the precision of the model across all larval labels (Figure 4.1D).

This classifier was then used to predict larval labels for adult clusters. Only neuronal clusters from the adult dataset were used for prediction. Importantly, the assignment of adult cells to larval clusters was agnostic to the adult cluster labels. After classifying each adult cell independently, we asked whether there was any correspondence between the original clusters of the adult dataset and those assigned by the GBT model. This result was presented in the form similar to a confusion matrix. We also verified that none of the cells mapped to the larval OB cluster because this region was actively dissected out in the single cell experiments performed with the adult telencephalon.

REFERENCES

Andero, R., Dias, B.G., and Ressler, K.J. (2014). A role for Tac2, NkB, and Nk3 receptor in normal and dysregulated fear memory consolidation. Neuron *83*, 444-454.

Griebel, G., Pichat, P., Boulay, D., Naimoli, V., Potestio, L., Featherstone, R., Sahni, S., Defex, H., Desvignes, C., Slowinski, F.*, et al.* (2016). The mGluR2 positive allosteric modulator, SAR218645, improves memory and attention deficits in translational models of cognitive symptoms associated with schizophrenia. Scientific reports *6*, 35320.

Li, M.L., Hu, X.Q., Li, F., and Gao, W.J. (2015). Perspectives on the mGluR2/3 agonists as a therapeutic target for schizophrenia: Still promising or a dead end? Prog Neuropsychopharmacol Biol Psychiatry *60*, 66-76.

Mossner, R., Schuhmacher, A., Schulze-Rauschenbach, S., Kuhn, K.U., Rujescu, D., Rietschel, M., Zobel, A., Franke, P., Wolwer, W., Gaebel, W.*, et al.* (2008). Further evidence for a functional role of the glutamate receptor gene GRM3 in schizophrenia. Eur Neuropsychopharmacol *18*, 768-772.

Neufeld-Cohen, A., Evans, A.K., Getselter, D., Spyroglou, A., Hill, A., Gil, S., Tsoory, M., Beuschlein, F., Lowry, C.A., Vale, W.*, et al.* (2010). Urocortin-1 and -2 double-deficient mice show robust anxiolytic phenotype and modified serotonergic activity in anxiety circuits. Molecular psychiatry *15*, 426-441, 339.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421-427.

Schwarzer, C. (2009). 30 years of dynorphins--new insights on their functions in neuropsychiatric diseases. Pharmacol Ther *123*, 353-370

Winham, S.J., Cuellar-Barboza, A.B., McElroy, S.L., Oliveros, A., Crow, S., Colby, C.L., Choi, D.S., Chauhan, M., Frye, M.A., and Biernacka, J.M. (2014). Bipolar disorder with comorbid binge eating history: a genome-wide association study implicates APOB. J Affect Disord *165*, 151-158.

Aoki, T., Kinoshita, M., Aoki, R., Agetsuma, M., Aizawa, H., Yamazaki, M., Takahoko, M., Amo, R., Arata, A., Higashijima, S*., et al.* (2013). Imaging of neural ensemble for the retrieval of a learned behavioral program. Neuron *78*, 881-894.

Biechl, D., Tietje, K., Ryu, S., Grothe, B., Gerlach, G., and Wullimann, M.F. (2017). Identification of accessory olfactory system and medial amygdala in the zebrafish. Sci Rep *7*, 44295.

Cheng, R.K., Jesuthasan, S.J., and Penney, T.B. (2014). Zebrafish forebrain and temporal conditioning. Philos Trans R Soc Lond B Biol Sci *369*, 20120462.

Dreosti, E., Lopes, G., Kampff, A.R., and Wilson, S.W. (2015). Development of social behavior in young zebrafish. Front Neural Circuits *9*, 39.

Ganz, J., Kroehne, V., Freudenreich, D., Machate, A., Geffarth, M., Braasch, I., Kaslin, J., and Brand, M. (2014). Subdivisions of the adult zebrafish pallium based on molecular marker analysis. F1000Res *3*, 308.

Griebel, G., Pichat, P., Boulay, D., Naimoli, V., Potestio, L., Featherstone, R., Sahni, S., Defex, H., Desvignes, C., Slowinski, F*., et al.* (2016). The mGluR2 positive allosteric modulator, SAR218645, improves memory and attention deficits in translational models of cognitive symptoms associated with schizophrenia. Scientific reports *6*, 35320.

J. Friedman (2001). Greedy Function Approximation: A Gradient Boosting Machine, . The Annals of Statistics *29*.

Lal, P., Tanabe, H., Suster, M.L., Ailani, D., Kotani, Y., Muto, A., Itoh, M., Iwasaki, M., Wada, H., Yaksi, E*., et al.* (2018). Identification of a neuronal population in the telencephalon essential for fear conditioning in zebrafish. BMC Biol *16*, 45.

Li, M.L., Hu, X.Q., Li, F., and Gao, W.J. (2015). Perspectives on the mGluR2/3 agonists as a therapeutic target for schizophrenia: Still promising or a dead end? Prog Neuropsychopharmacol Biol Psychiatry *60*, 66-76.

Mossner, R., Schuhmacher, A., Schulze-Rauschenbach, S., Kuhn, K.U., Rujescu, D., Rietschel, M., Zobel, A., Franke, P., Wolwer, W., Gaebel, W.*, et al.* (2008). Further evidence for a functional role of the glutamate receptor gene GRM3 in schizophrenia. Eur Neuropsychopharmacol *18*, 768-772.

Neufeld-Cohen, A., Evans, A.K., Getselter, D., Spyroglou, A., Hill, A., Gil, S., Tsoory, M., Beuschlein, F., Lowry, C.A., Vale, W.*, et al.* (2010). Urocortin-1 and -2 double-deficient mice show robust anxiolytic phenotype and modified serotonergic activity in anxiety circuits. Molecular psychiatry *15*, 426-441, 339.

Pandey, S., Shekhar, K., Regev, A., and Schier, A.F. (2018). Comprehensive Identification and Spatial Mapping of Habenular Neuronal Types Using Single-Cell RNA-Seq. Curr Biol *28*, 1052-1065 e1057.

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

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421-427.

Stednitz, S.J., McDermott, E.M., Ncube, D., Tallafuss, A., Eisen, J.S., and Washbourne, P. (2018). Forebrain Control of Behaviorally Driven Social Orienting in Zebrafish. Curr Biol *28*, 2445-2451 e2443.

Thyme, S.B., Pieper, L.M., Li, E.H., Pandey, S., Wang, Y., Morris, N.S., Sha, C., Choi, J.W., Herrera, K.J., Soucy, E.R.*, et al.* (2019). Phenotypic Landscape of Schizophrenia-Associated Genes Defines Candidates and Their Shared Functions. Cell *177*, 478-491 e420.

Waltman, L., and Eck, N.J.v. (2013). A smart local moving algorithm for large-scale modularity-based community detection. Eur Phys J B *86*, 1–33.

Winham, S.J., Cuellar-Barboza, A.B., McElroy, S.L., Oliveros, A., Crow, S., Colby, C.L., Choi, D.S., Chauhan, M., Frye, M.A., and Biernacka, J.M. (2014). Bipolar disorder with comorbid binge eating history: a genome-wide association study implicates APOB. J Affect Disord *165*, 151-158.

Zarate, Y.A., and Fish, J.L. (2017). SATB2-associated syndrome: Mechanisms, phenotype, and practical recommendations. Am J Med Genet A *173*, 327-337.

Zeisel, A., Hochgerner, H., Lonnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Haring, M., Braun, E., Borm, L.E., La Manno, G.*, et al.* (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999-1014 e1022.

Chapter 5: Conclusions and prospects

Through my graduate work, we have uncovered a comprehensive cell type landscape of two regions of the zebrafish brain: habenula and telencephalon. The cell type landscape of the habenula presented a blueprint for comprehensive discovery of rare neuronal types and integrative analysis of molecular, spatial and functional properties (Chapter 2 and 3). In generating the cell type atlas of the habenula, we presented the following advances. First, we devised a robust protocol for the dissociation and capture of single neurons from the zebrafish brain. Second, we found that comprehensive identification of neuronal types by scRNA-seq can be achieved by high cell-sampling coverage of a small brain region. Third, we discovered thirteen new neuronal types, identified fine-grained spatial subdivisions in the habenula and defined dozens of new marker genes for novel habenular subtypes. Surprising, we found that diverse neuronal types are largely retained from larva to adult (Chapter 2). Fourth, we generated a spatial *in situ* atlas of the habenula using computational image registration. Using this atlas and immediate early gene *cfos*, I also identified a cellular population in the ventro-lateral habenula that mediates behavioral responses to inescapable aversive environmental stimuli (Chapter 3). Through this work, I also created a resource for future studies on habenular development and function and provided a technological framework for the characterization of other brain regions.

I hope our work on the habenular neuronal types will spur future studies on the role of these cell types and marker genes in generating habenular behaviors. Our work of mapping neuronal activity onto single-cell RNAseq clusters using the *in situ* atlas established the ventrolateral habenular cell type as an important node for mediating behavioral responses to inescapable aversive stimuli. Since then, additional studies based on calcium imaging have demonstrated the dynamic role that this ventro-lateral cell type may play in mediating behavioral responses to inescapable aversive stimuli (Andalman et al., 2019; Rodriguez-Sosa et al., 2019).

Furthermore, using an oversampling approach with scRNAseq, we identified rare neuronal populations that were represented by less than 20 neurons *in vivo*. Marker genes for these neuronal populations can now be used to interrogate the function of these small cell groups that had so far evaded functional studies due to a lack of specific genes that distinguish these cell populations from others. Indeed, recent studies have found that Hb03, one of the smaller clusters in our data marked by *spx* is involved in the regulation of anxiety and that specific over-expression of the marker gene *spx* within the habenula reduces anxiety in zebrafish (Jeong et al., 2019).

5.1 The Role of Ventral Habenular Cell Type-specific Genes in Regulating Locomotor Activity

I followed up on the habenular atlas by interrogating the role of cell type specific marker genes in generating habenular function (Chapter 3). After generating CRISPR mutants of genes specific to the ventral habenular clusters, I measured the baseline brain activity and behavioral phenotypes of these mutants. Strikingly, all the genes mutated from ventral habenular clusters had mild to strong phenotypes associated with changes in locomotor activity during the day. This raises interesting questions that I hope will be explored in the future.

Among the four genes, *aoc1* which encodes for a histamine deaminase when knocked out in zebrafish results in both a reduction in arousal threshold and an increase in locomotor activity (Figure 3.6). While the increase in locomotor activity occurred only during the day, the reduction in arousal threshold was observed both during day and night. It has long been known that habenula exerts a regulatory role on sleep and that hyperactivity within the lateral habenula may be the link between depression and sleep disturbance (Aizawa et al., 2013). It is thought that this regulation occurs through the habenular modulation of the serotonergic system. Our results indicate that there might be additional modes by which the habenula might regulate sleep. For instance, *aoc1* mutants display increased brain activity in the ventral habenula, the homolog of mammalian lateral habenula. Since histamine release promotes wakefulness, it is not unlikely that the specific expression of *aoc1* in the ventral habenula might be one of the other ways in which habenula may regulate sleep i.e. through the regulation of histamine. Further experimentation will be required to interrogate if the mutants have altered histamine levels.

Another interesting line of inquiry would be to test the behavioral strategies of these ventral habenular mutants when confronted with aversive environmental stimuli. While our studies with *cfos* and others with calcium imaging have established the ventral habenula as an important node for generating behavioral responses to aversive environmental conditions, it is unclear if the top marker genes of ventral habenular types are also important for these behavioral responses. This represents an excellent opportunity to ask whether there is harmony between markers found by single cell analysis and the functional properties of a particular cell type within a circuit that modulates certain behaviors. There are a number of inescapable aversive stimuli that could be presented to mutant and wild type larvae in a high throughput manner, including heat and high salt conditions. Of the four genes, it would be particularly interesting if *gpr139* mutants display different coping strategies to these aversive conditions. Consistent with the hyperactivation of lateral habenula in the pathophysiology of depression, *gpr139-/-* animals display change in brain activity phenotype in the ventral habenula. Furthermore, *gpr139* has been implicated in depression by GWAS (Dunn et al., 2016). Therefore, the modulation of receptor activity of *gpr139* may represent an exciting opportunity for the treatment of disorders such as depression.

5.2 Conservation of Habenular Cell types Across Different Species

It will also be interesting to follow up the study of the habenula in fish with transcriptional profiling of habenula in other species. An important debate in the cell type classification field has been whether or not to harmonize cell type definitions across multiple species.

Habenula is one of the most conserved brain regions that is present within all vertebrate species (Hikosaka, 2010). Previous studies suggest that homologous anatomical and molecular features in the habenula exist between mice and fish (Amo et al., 2010; Bianco and Wilson, 2009; Gamse et al., 2005; Stephenson-Jones et al., 2012). The shared features in cytoarchitecture, projection patterns and functions between habenula of different species has been extensively studied. The comprehensive profiling of cell types within the zebrafish habenula allows us to ask questions about the conservation of these cell types and molecular profiles within the mammalian habenula. Indeed, a version of this analysis was performed recently by a study that explored the homology of transcriptional profiles within the mouse habenula to those found by our study in the fish habenula (Yoshiko Hashikawa1, 2019).

5.3 Matching Molecular Properties with Neuronal Dynamics and Connectivity

Another interesting line of study would be to match the molecular profiles of habenular neurons found by single cell RNAseq to physiological/firing properties of neurons found by calcium imaging or electrophysiological measurements. Do transcriptionally distinct habenular neurons also have different neuronal firing dynamics? We attempted to follow this line of inquiry by performing calcium imaging in the habenula in response to aversive stimulus and generating a map of the neuronal dynamics with the goal of interfacing it with molecular atlas (Figure A.9). We found a diversity of different neuronal types within the habenula spread along the dorso-ventral axis. However, it was difficult interface the maps obtained by calcium imaging to those obtained via in situ hybridization of single genes. In future, this question can be systematically probed by other approaches. Techniques such as Patch-seq allow simultaneous measurement of electrophysiological and molecular properties of single neurons (van den Hurk et al., 2018). Similarly, multiplexed in situ approaches such as MERFISH or seq-FISH in which many genes can be probed at the same time can be used to interface calcium imaging maps with in situ approaches. Our study also raises the possibility of a finer topography of efferent projections of transcriptionally distinct neuronal subsets into finer sub-regions of downstream targets. Another potential future study might be to generate transgenic reporter lines for the marker genes found by single cell analysis to ask if the distinct subsets might innervate distinct sub-regions of downstream IPN or Raphe nucleus. Indeed, a single cell study in the mouse habenula has since then shown that the distinct neuronal subtypes within the lateral habenula innervate distinct regions of the ventral tegmental area (Wallace, 2019).

5.4 Cell Type Dissection of Behavior and Brain Activity Phenotypes using scRNAseq

Our work on the telencephalon represents a way in which brain activity and behavioral phenotypes can be dissected at a molecular level in a cell type specific manner using scRNAseq. We first generated a comprehensive survey of cell types in the telencephalon in both larvae and adult. In contrast to the habenula, we found that the telencephalon established new cell type domains going from a larva to an adult. We then surveyed the telencephalic cell types of a *znf536* mutant to discovery loss of specific neuropeptidergic cell types within the pallium as well as the subpallium. This result raises interesting questions that can be answered in the future. First, *znf536* is widely expressed in the forebrain across many cell types. Why does the loss of such a broadly expressed transcription factor lead to effects on small subpopulations of peptidergic neurons? Future studies might find binding sites for *znf536* as well as combine cell type profiling with single cell ATAC-seq to define regions of open chromatin in affected and non-affected cell types. This might help us understand the specific vulnerabilities of certain cell types to the loss of *znf536*. Second, the loss of these neurons was accompanied by an increase in the number of immature neurons in *znf536* mutants, suggesting that the phenotype may be a result of a maturation defect. In this case, it will be interesting to compare pseudolineage trees of mutant animals with wild type animals to uncover how the gene expression cascades might differ in mutant and wildtype animals along different stages of maturation. Our studies provide the framework for answering these questions in the future.

REFERENCES

Aizawa, H., Cui, W., Tanaka, K., and Okamoto, H. (2013). Hyperactivation of the habenula as a link between depression and sleep disturbance. Front Hum Neurosci *7*, 826.

Amo, R., Aizawa, H., Takahoko, M., Kobayashi, M., Takahashi, R., Aoki, T., and Okamoto, H. (2010). Identification of the zebrafish ventral habenula as a homolog of the mammalian lateral habenula. J Neurosci *30*, 1566-1574.

Andalman, A.S., Burns, V.M., Lovett-Barron, M., Broxton, M., Poole, B., Yang, S.J., Grosenick, L., Lerner, T.N., Chen, R., Benster, T., *et al.* (2019). Neuronal Dynamics Regulating Brain and Behavioral State Transitions. Cell *177*, 970-985 e920.

Bianco, I.H., and Wilson, S.W. (2009). The habenular nuclei: a conserved asymmetric relay station in the vertebrate brain. Philos Trans R Soc Lond B Biol Sci *364*, 1005-1020.

Dunn, E.C., Wiste, A., Radmanesh, F., Almli, L.M., Gogarten, S.M., Sofer, T., Faul, J.D., Kardia, S.L., Smith, J.A., Weir, D.R., *et al.* (2016). Genome-Wide Association Study (Gwas) and Genome-Wide by Environment Interaction Study (Gweis) of Depressive Symptoms in African American and Hispanic/Latina Women. Depress Anxiety *33*, 265-280.

Gamse, J.T., Kuan, Y.S., Macurak, M., Brosamle, C., Thisse, B., Thisse, C., and Halpern, M.E. (2005). Directional asymmetry of the zebrafish epithalamus guides dorsoventral innervation of the midbrain target. Development *132*, 4869-4881.

Hikosaka, O. (2010). The habenula: from stress evasion to value-based decision-making. Nat Rev Neurosci *11*, 503-513.

Jeong, I., Kim, E., Seong, J.Y., and Park, H.C. (2019). Overexpression of Spexin 1 in the Dorsal Habenula Reduces Anxiety in Zebrafish. Front Neural Circuits *13*, 53.

Rodriguez-Sosa, N., Biswas, S., and Shabel, S.J. (2019). Coping with Stress, One Habenula Neuron at a Time. Neuron *102*, 520-522.

Stephenson-Jones, M., Floros, O., Robertson, B., and Grillner, S. (2012). Evolutionary conservation of the habenular nuclei and their circuitry controlling the dopamine and 5-hydroxytryptophan (5-HT) systems. Proc Natl Acad Sci U S A *109*, E164-173.

van den Hurk, M., Erwin, J.A., Yeo, G.W., Gage, F.H., and Bardy, C. (2018). Patch-Seq Protocol to Analyze the Electrophysiology, Morphology and Transcriptome of Whole Single Neurons Derived From Human Pluripotent Stem Cells. Front Mol Neurosci *11*, 261.

Wallace, K.W.H., Daniel Hochbaum, Minsuk Hyun, Gianna Radeljic, Bernardo L. Sabatini (2019). Distinct neuronal subtypes of the lateral habenula differentially target ventral tegmental area dopamine neurons. Biorxiv.

Yoshiko Hashikawa1, K.H., Marcus L. Basiri, Yuejia Liu, Nathan L. Johnston, Omar R. Ahmad, Garret D. Stuber (2019). Transcriptional and Spatial Resolution of Cell Types in the Mammalian Habenula. Biorxiv

Appendix

Appendix 1: Supplementary Figures for Chapter 2

Figure A.1. Quality Control of Larval Droplet Data.

(A) Amount of RNA extracted (in ng) from variable numbers of FAC-sorted mouse dendritic cells of the immune system and larval zebrafish neurons.

(B) Estimated average RNA amount per cell (in pg) in zebrafish neurons and mouse dendritic cells.

(C) Distribution of genes detected across the filtered droplet larval dataset.

(D) Distribution of UMIs detected across the filtered droplet larval dataset.

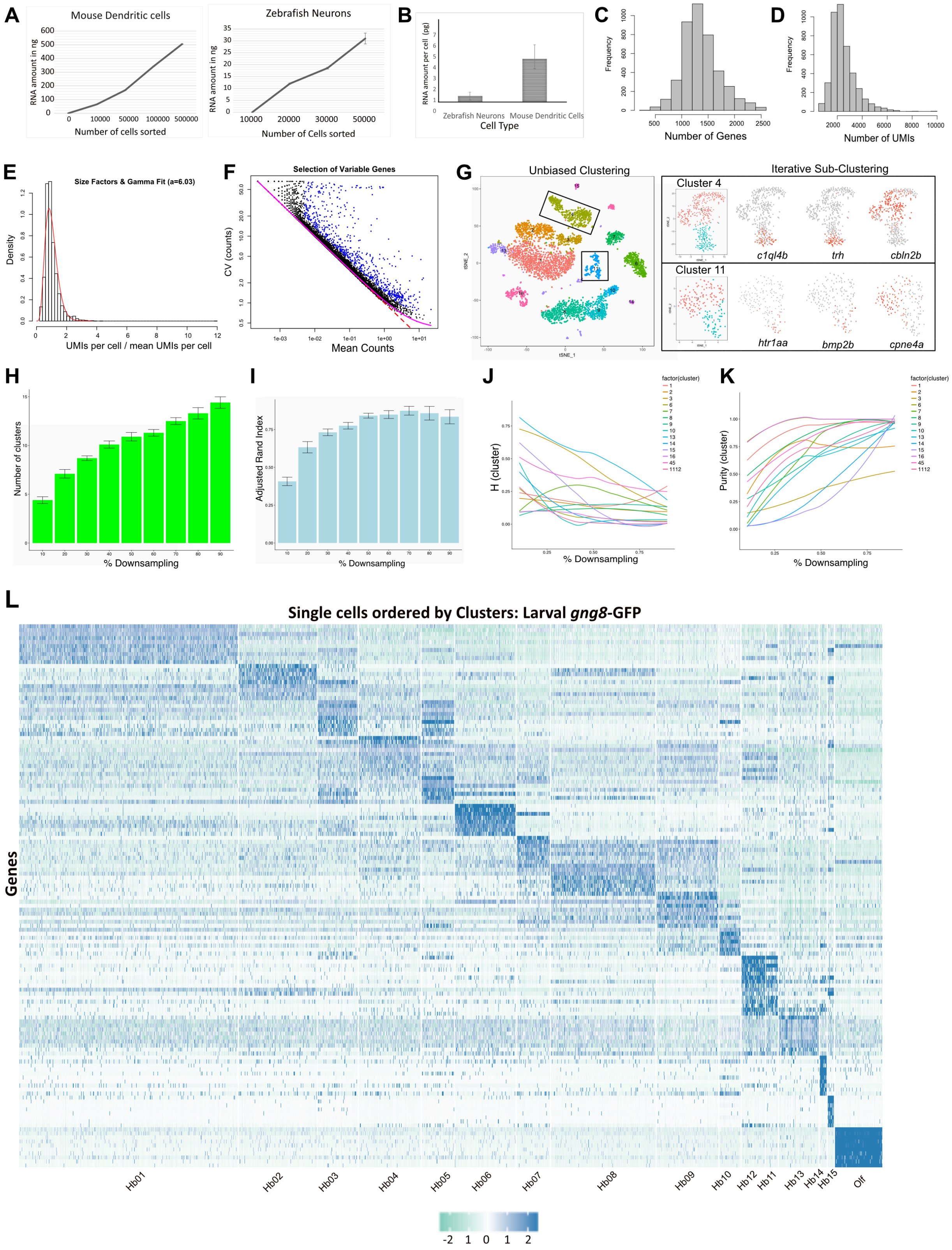
(E) The distribution of the relative library sizes η of cells in droplet data modeled as a Gamma distribution (red curve). Note that our particular parameterization of η, ensures that the Gamma distribution has mean of 1, making it a single parameter fit.

(F) Plot of CV (coefficient of variation, i.e. ratio of the standard deviation and the mean) of UMI counts across cells versus mean UMI counts. Each black dot corresponds to a single gene; Red dotted line is the empirical mean-CV relation predicted by the Poisson model; the magenta curve represents our modified Gamma-Poisson (or Negative Binomial) model of the expected null CV as a function of mean counts.

(G) tSNE visualization of the clusters obtained from the larval droplet dataset with unbiased clustering (left). All clusters were subjected to iterative sub-clustering. Iterative Sub-clustering of Cluster 4 and Cluster 11 separate them into smaller sub-clusters (right) each distinguished by the expression of a unique set of genes shown in panels in the right.

(H-I) Downsampling experiments to evaluate robustness of the droplet larval clustering. The automated clustering pipeline used in these experiments is the same as was used for the full dataset. For consistency, cluster pairs 4 and 5, and 11 and 12 in the full dataset are “merged” in these comparisons, since these were originally split through *post hoc* analyses. Results represent ten independent trials conducted at each level of downsampling (10% - 90% of the full dataset). H) Number of clusters at different levels of downsampling. I) Adjusted rand index (ARI) comparing the consistency with the original clustering. Trials with > 50% of the dataset produce high values of ARI > 0.85. J) The normalized Shannon entropy (H) of each original cluster in the downsampled dataset. $H_{(cluster)}$ is a measure of the tightness with which members of an original cluster are grouped together in the downsampled data, with 0 indicating perfect clustering. As expected, a majority of clusters exhibit low values of H (<0.25) as the sample size increases. The entropy of cluster 1 exhibits an increase towards the end as we noticed that the automatic algorithm splits this cluster in some trials at >80% downsampling. These sub-clusters, however, did not show sufficient gene expression differences justifying a subdivision within our dataset. K) The Cluster purity (P) of each cluster, indicating how exclusively members of an original cluster are placed together, avoiding members of other clusters. Most clusters exhibit a monotonic increase in purity with larger sample sizes. The only exceptions are clusters 2 and 3, which appear to slightly intermix at all levels because they are closely related.

(L) Heatmap of relative gene expression showing differentially expressed genes across the 15 habenular neuronal types and olfactory placode cluster. Rows correspond to individual genes enriched in individual clusters based on a bimodal test (McDavid et al., 2012), and columns are individual cells, ordered by cluster identity shown on the t-SNE Plot in Figure 1D.


A
Mouse Dendritic cells
RNA amount in ng
Number of cells sorted
Zebrafish Neurons
RNA amount in ng
Number of Cells sorted
B
RNA amount per cell (pg)
Zebrafish Neurons
Mouse Dendritic Cells
Cell Type
C
Frequency
Number of Genes
D
Frequency
Number of UMIs
E
Size Factors & Gamma Fit (a=6.03)
Density
UMIs per cell / mean UMIs per cell
F
Selection of Variable Genes
CV (counts)
Mean Counts
G
Unbiased Clustering
Iterative Sub-Clustering
Cluster 4
c1ql4b
trh
cbln2b
Cluster 11
htr1aa
bmp2b
cpne4a
tSNE_1
tSNE_2
H
Number of clusters
% Downsampling
I
Adjusted Rand Index
% Downsampling
J
H (cluster)
% Downsampling
factor(cluster)
K
Purity (cluster)
% Downsampling
factor(cluster)
L
Single cells ordered by Clusters: Larval gng8-GFP
Genes
Hb01
Hb02
Hb03
Hb04
Hb05
Hb06
Hb07
Hb08
Hb09
Hb10
Hb12
Hb11
Hb13
Hb14
Hb15
Olf
-2
1
0
1
2

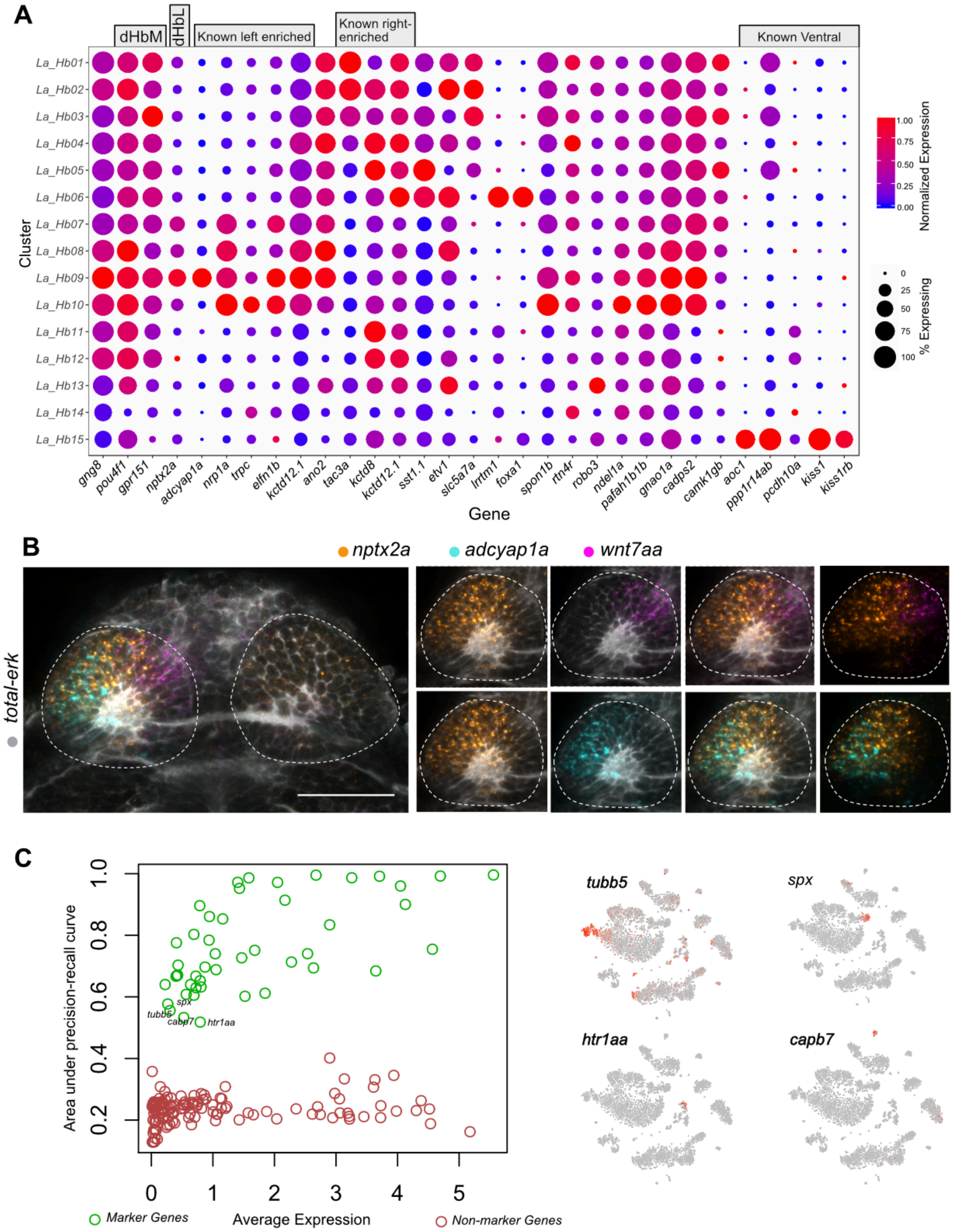


Figure A.2. Mapping of Previously Characterized Regionalized Genes onto Droplet Clusters.

(A) Expression patterns of previously described regionalized habenular genes. Genes that were previously described to be left-right asymmetric and specific to sub-regions (dHbM = dorso-medial habenula, dHbL = dorso-lateral habenula, vHb = ventral habenula) are highlighted with a grey box on top.

Figure A.2 (Continued). (B) *In vivo* expression pattern of *nptx2a* (dHbL marker) together with registered RNA-FISH signals of marker genes. The largest overlap of high *nptx2a* expressing cells is with the left enriched clusters, Hb08 (*wnt7aa*$^+$) and Hb09 (*adcyap1a*$^+$). Panels on the right show individual markers in the left habenula. Scale bar indicates 50 μm.

(C) Quantification of specificity of marker genes found by single cell analysis. Area under the precision recall curve (AUCPR) for marker genes (green) versus non-marker genes(red) within the same range of expression value [Right panel]. A few marker genes with low AUCPR values constitute those that are expressed in smaller clusters or that belong to the immature neuron cluster (Hb13) whose signatures are spread across multiple clusters in the dataset (see t-SNE plots in the right).

Figure A.3. Spatial Registration and Whole Brain Expression Pattern of Habenular Markers, Related to Figures 2.2 and 2.3.

(A) Total-Erk (t-Erk) confocal stacks of the habenula registered to a reference stack using Computational Morphometry Toolkit (CMTK). Shown are outputs across three fish (magenta, yellow and cyan).

Figure A.3 (Continued). (B) t-Erk confocal stacks of the whole brain registered to a reference stack through using Advanced Normalization Tools (ANTS) software package. Shown are outputs across three fish (magenta, yellow and cyan).

(C) Maximum intensity projections of whole brain expression patterns of key habenular markers validated in this study to obtain a list of markers that are specific to habenular neuronal types across the brain. Full stacks are available through a linked website [See Additional Resources Section].

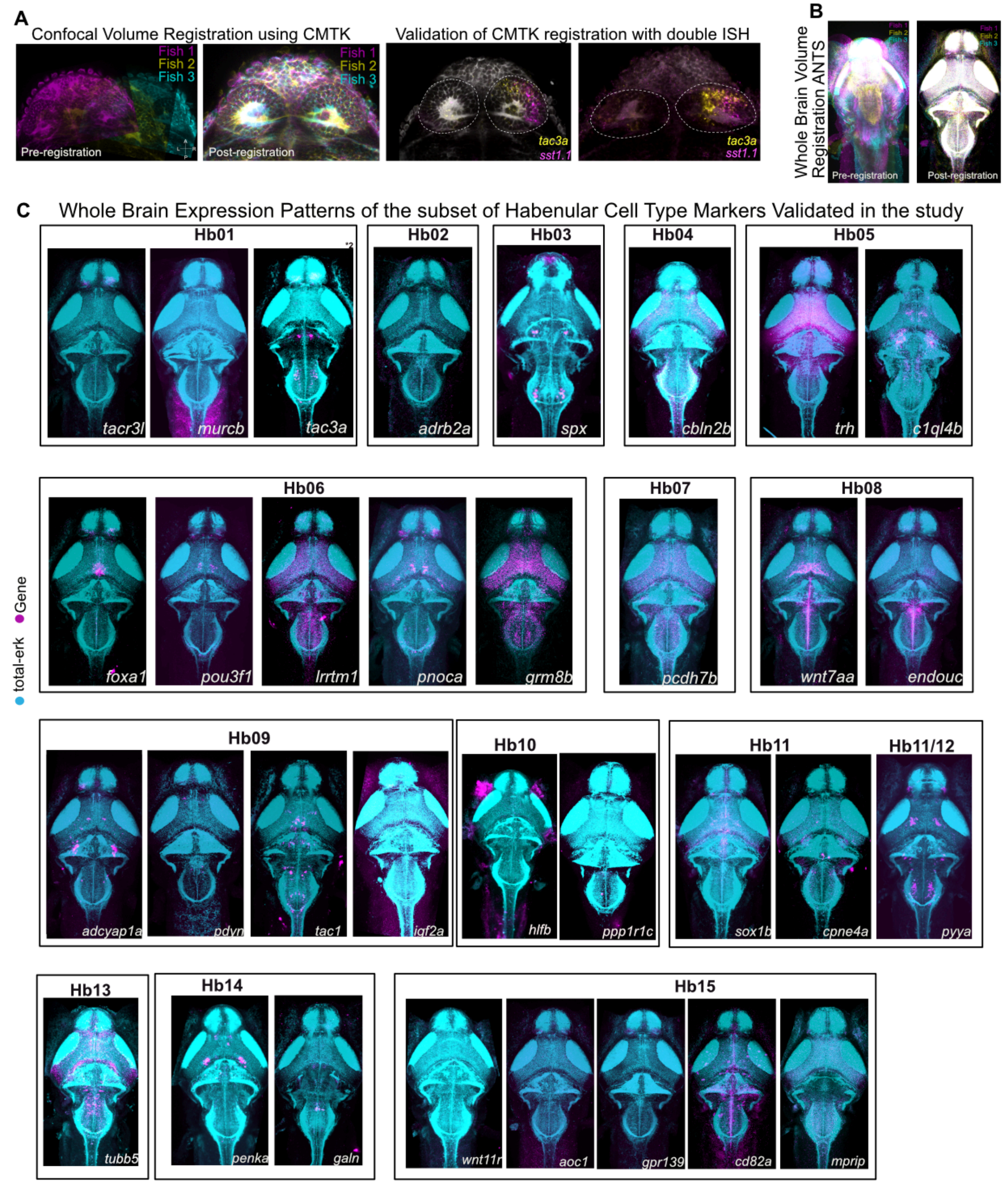

A
Confocal Volume Registration using CMTK
Fish 1
Fish 2
Fish 3
Pre-registration
Post-registration
Validation of CMTK registration with double ISH
tac3a
sst1.1
B
Whole Brain Volume Registration ANTS
Fish 1
Fish 2
Fish 3
Pre-registration
Post-registration
C
Whole Brain Expression Patterns of the subset of Habenular Cell Type Markers Validated in the study
total-erk
Gene
Hb01
tacr3l
murcb
tac3a
Hb02
adrb2a
Hb03
spx
Hb04
cbln2b
Hb05
trh
c1ql4b
Hb06
foxa1
pou3f1
lrrtm1
pnoca
grm8b
Hb07
pcdh7b
Hb08
wnt7aa
endouc
Hb09
adcyap1a
pdyn
tac1
igf2a
Hb10
hlfb
ppp1r1c
Hb11
sox1b
cpne4a
Hb11/12
pyya
Hb13
tubb5
Hb14
penka
galn
Hb15
wnt11r
aoc1
gpr139
cd82a
mprip

Figure A.4. Quality Control Metrics of SMART-seq2 Data, Related to Figure 2.4.

(A) Distribution of number of genes detected per cell across the filtered SMART-seq2 dataset

(B) Distribution of read counts per cell across the filtered SMART-seq2 dataset.

(C) Composition of each cluster by experimental batch. Batch1 represents cells from an enhancer trap line that labels a subset of cells labeled by the *gng8*-GFP line. Batch 2 and Batch 3 represent two replicates of the *gng8*-GFP transgenic line.

(D) Gamma distribution fit (blue curve) of the distribution of size-factors for SMART-seq2 dataset.

(E) CV versus mean transcript counts of all expressed genes in the SMART-seq2 dataset together with the theoretically expected Poisson relationship (dotted red line) and the Poisson Gamma relationship (magenta solid line). The *Brennecke et al.* fit is shown in yellow.

(F) CV versus mean read counts of the larval droplet data with Poisson (dotted red line), Poisson Gamma (magenta solid line) and *Brennecke* et al. fits (yellow solid line).

(G) CV vs mean read counts of mouse genes in the Tasic, Menon et al. dataset (GSE71585) together with Poisson, Poisson-Gamma relationship (magenta solid line) and *Brennecke* et al. fits (yellow solid line).

(H) CV versus mean transcript counts of the retinal bipolar neuron Drop-seq data with Poisson (dotted red line), Poisson Gamma (magenta solid line) and *Brennecke* et al. fits (yellow solid line).

(I) Performance of random forest (RF) model trained on the larval droplet dataset with the graph clustering labels shown in Figure 1C. A training set was formed by choosing 70% of the cells from the entire dataset, with proportional representation from each cluster. The trained RF model was then used to classify each cell in the remaining 30% of the data (test set) into learned cluster labels. The result of the classification of the test set is shown as a confusion matrix. The last column represents cells that did not receive a maximum vote by a margin of > 15 % of the trees, and hence were labeled “unassigned”.

(J) Correlation of the average expression levels of genes in the SS2 dataset (x-axis) versus the proportion of cells expressing each gene (y-axis).

(K) Correlation of the average expression levels of genes (x-axis) in the droplet dataset versus the proportion of cells expressing each gene (y-axis).

(L) Scatter plot of the proportion of cells expressing each gene in droplet data (y-axis) versus the similar proportion in the SS2 data (x-axis).

(M) Comparison of the distribution of Shannon entropy for two sets of genes in the droplet dataset.1) Top 25 marker genes found through independent analysis of SS2 dataset (blue), 2) the set of genes that were expressed in high proportion of cells in the SS2 data and had little to no expression in the droplet dataset (red curve, genes labeled in red in Figure S4J).

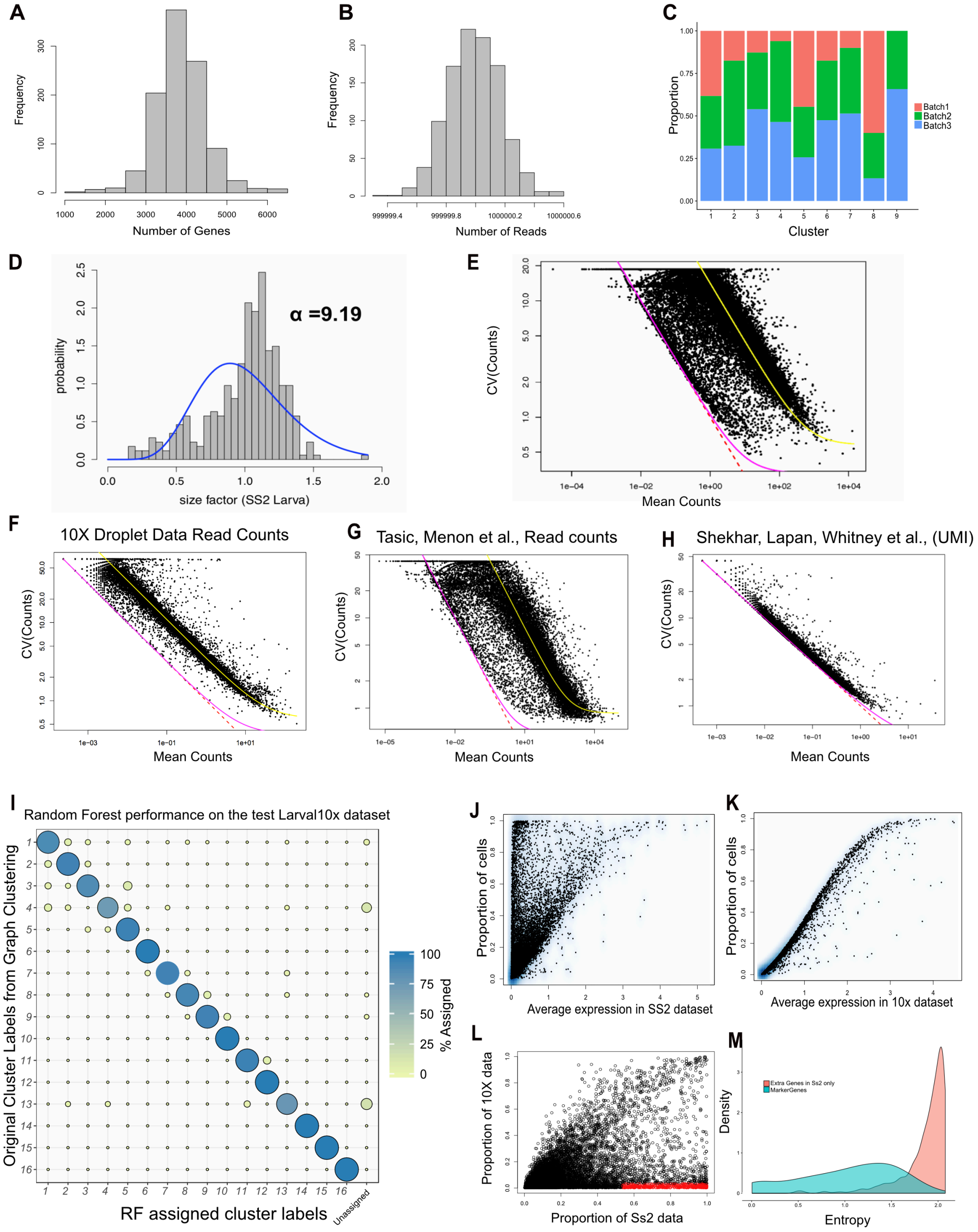

A
Frequency
Number of Genes
B
Frequency
Number of Reads
C
Proportion
Cluster
Batch1
Batch2
Batch3
D
probability
size factor (SS2 Larva)
α =9.19
E
CV(Counts)
Mean Counts
F
10X Droplet Data Read Counts
CV(Counts)
Mean Counts
G
Tasic, Menon et al., Read counts
CV(Counts)
Mean Counts
H
Shekhar, Lapan, Whitney et al., (UMI)
CV(Counts)
Mean Counts
I
Random Forest performance on the test Larval10x dataset
Original Cluster Labels from Graph Clustering
RF assigned cluster labels
Unassigned
% Assigned
J
Proportion of cells
Average expression in SS2 dataset
K
Proportion of cells
Average expression in 10x dataset
L
Proportion of 10X data
Proportion of Ss2 data
M
Density
Entropy
Extra Genes in Ss2 only
MarkerGenes

Figure A.5. Quality Control of Adult Single Cell RNA-seq Dataset, Related to Figure 5.

(A) Distribution of genes detected across the filtered adult droplet dataset.

(B) Distribution of UMIs detected across the filtered adult droplet dataset.

(C) Composition of each cluster by experimental batch post batch correction and filtering.

(D) FISH for ventral cluster-specific genes *cd82a* and *mprip* that are also identified to be expressed in a regional manner in the larval ventral habenula.

(E) FISH labeling of *fabp7a*, marker for a progenitor-like cells in the adult and larval habenula.

(F) Violin plot showing expression of *gng8* in larval and adult dataset. The expression values of *gng8* (log(TPM+1) >2) were used to filter the adult dataset for RF analysis. The right panel shows a t-SNE plot of the resulting *gng8* positive in the adult dataset.

(G) Violin plots showing examples of genes differentially expressed between Ad_Hb02A and Ad_Hb02B. *Rac2* is enriched in Ad_Hb02A(right) and *nebl* is enriched in Ad_Hb02B(left) (upper panel). In situ hybridization showing right-enriched expression of *nebl* in the adult habenula.

(H) Schematic showing the validated neuronal types in the adult habenula in a single coronal slice. Schematic represents a projection across all slices assayed. Only information about left–right and dorso-ventral positioning of neuronal types is represented.

(I) Anatomical reorganization of the habenula between the time points analyzed in the study and the rough mapping of select neuronal types onto this map. As described by previous literature, the habenula undergoes complex morphogenetic changes in which the dorso-medial region migrates outwards laterally and the ventrolateral region migrates inwards medially. The positioning of the neuronal types found in our study corroborates this migratory pattern.

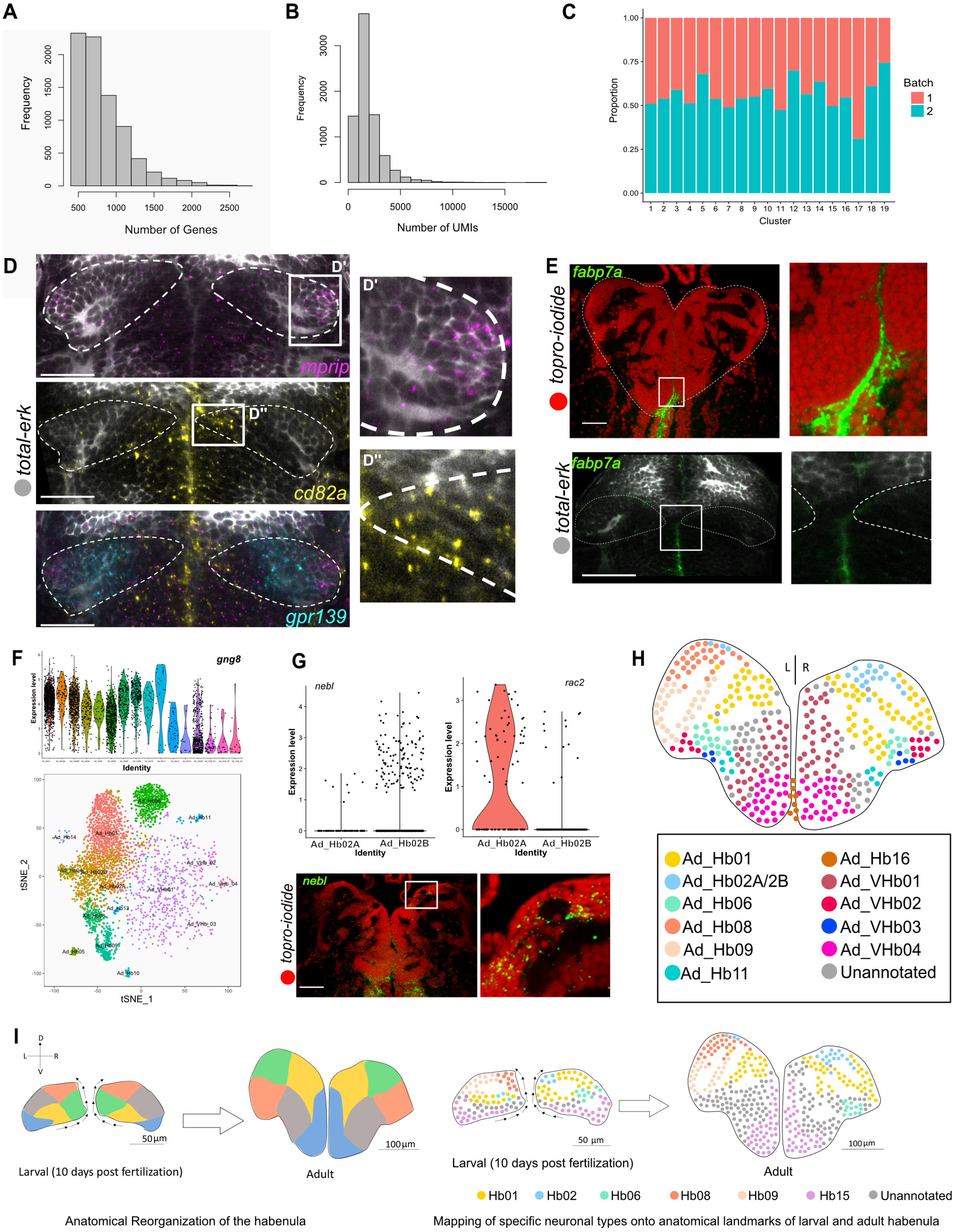

A
Frequency
Number of Genes
0
500
1000
1500
2000
2500
B
Frequency
Number of UMIs
0
1000
2000
3000
5000
10000
15000
C
Proportion
Cluster
0.00
0.25
0.50
0.75
1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Batch
1
2
D
D'
D''
total-erk
mprip
cd82a
gpr139
E
fabp7a
topro-iodide
fabp7a
total-erk
F
gng8
Expression level
Identity
tSNE_2
tSNE_1
G
nebl
rac2
Expression level
Ad_Hb02A
Ad_Hb02B
Identity
nebl
topro-iodide
H
L
R
Ad_Hb01
Ad_Hb02A/2B
Ad_Hb06
Ad_Hb08
Ad_Hb09
Ad_Hb11
Ad_Hb16
Ad_VHb01
Ad_VHb02
Ad_VHb03
Ad_VHb04
Unannotated
I
D
L
R
V
50 µm
Larval (10 days post fertilization)
100 µm
Adult
50 µm
Larval (10 days post fertilization)
100 µm
Adult
Hb01
Hb02
Hb06
Hb08
Hb09
Hb15
Unannotated
Anatomical Reorganization of the habenula
Mapping of specific neuronal types onto anatomical landmarks of larval and adult habenula

Table A.1. Curated Markers for the larval habenula.

p_Val	Avg_diff	Cluster	Gene	Zfin accession
0	2.1963462	Hb01	*g0s2*	ZDB-GENE-081022-103
0	1.95089206	Hb01	*adcyap1b*	ZDB-GENE-020809-4
0	1.56182068	Hb01	*tacr3l*	ZDB-GENE-061207-19
3.33e-314	1.87393612	Hb01	*fxyd1*	ZDB-GENE-050309-14
2.32e-310	1.57555627	Hb01	*murcb*	ZDB-GENE-041212-87
3.97E-273	1.33451062	Hb01	*prox1a*	ZDB-GENE-980526-397
5.68E-180	1.2915815	Hb01	*impact*	ZDB-GENE-040927-10
3.37E-265	1.55420984	Hb01	*nrgna*	ZDB-GENE-090710-4
1.21E-135	1.14665915	Hb01	*ttna*	ZDB-GENE-030113-2
1.60E-229	1.27062108	Hb01	*kcnd1*	ZDB-GENE-081105-40
6.16E-198	1.4024978	Hb01	*sdc4*	ZDB-GENE-061111-1
4.46E-185	1.5563455	Hb02	*adrb2a*	ZDB-GENE-100414-3
4.79E-98	1.12566886	Hb02	*vav3b*	ZDB-GENE-070912-251
4.62E-110	1.58631704	Hb02	*gng2*	ZDB-GENE-050417-59
7.44E-102	1.23267433	Hb02	*rgma*	ZDB-GENE-040527-1
9.21E-102	0.9832537	Hb02	*cyth4a*	ZDB-GENE-090624-6
1.67E-78	2.36732895	Hb03	*spx*	ZDB-GENE-041210-158
1.58E-62	2.09216043	Hb03	*spa17*	ZDB-GENE-061027-337
2.28E-195	2.359403	Hb04	*cbln2b*	ZDB-GENE-070615-12
2.23E-104	0.7233532	Hb04	*dcdc2c*	
5.26E-299	1.2289715	Hb05	*c1ql4b*	ZDB-GENE-131127-59
7.83E-220	1.040571	Hb05	*c1ql4a*	ZDB-GENE-090313-185
5.121743e	1.9662987	Hb05	*trh*	ZDB-GENE-020930-1
2.43E-87	0.5764724	Hb05	*ghra*	ZDB-GENE-070509-1
0	2.30213246	Hb06	*foxa1*	ZDB-GENE-990415-78
1.09e-311	2.3294158	Hb06	*lrrtm1*	ZDB-GENE-050506-80
1.73E-258	2.26192052	Hb06	*pou3f1*	ZDB-GENE-980526-372
2.36E-203	2.03746148	Hb06	*pnoca*	ZDB-GENE-101229-1
1.15E-140	0.9667398	Hb06	*grm8b*	ZDB-GENE-110421-3
3.17E-192	1.94459897	Hb06	*im:7152348*	ZDB-GENE-041111-308
1.31E-189	1.85119093	Hb06	*pmp22b*	ZDB-GENE-060421-4337
8.90E-78	1.0310556	Hb06	*gpr78*	ZDB-GENE-110411-251
2.45E-109	0.8582727	Hb06	*oprl1*	ZDB-GENE-040312-4
9.56E-101	1.61846599	Hb07	*pcdh7b*	ZDB-GENE-081104-299
4.93E-82	1.39875526	Hb07	*spon1a*	ZDB-GENE-000427-9
2.02E-72	1.30988217	Hb07	*tmem98*	ZDB-GENE-070112-1092
5.08E-62	1.41916147	Hb07	*eno1b*	ZDB-GENE-040426-1651
2.65E-212	1.41769201	Hb08	*wnt7aa*	ZDB-GENE-051129-1
8.32E-202	1.23412746	Hb08	*pth2r*	ZDB-GENE-991123-10
4.49E-141	1.37118246	Hb08	*endouc*	ZDB-GENE-060503-141
5.90E-268	3.1794816	Hb09	*adcyap1a*	ZDB-GENE-020809-4
2.83E-169	1.74269084	Hb09	*igf2a*	ZDB-GENE-991111-3

Table A.1 (Continued).

3.83E-122	1.43998887	Hb09	*pdyn*	ZDB-GENE-060417-1
4.48E-297	1.102586	Hb09	*tacr2*	ZDB-GENE-090312-188
1.85E-235	1.061916	Hb09	*adra1d*	ZDB-GENE-090312-203
1.38E-115	1.54959126	Hb09	*bx957297.1*	
2.94E-97	1.48947036	Hb09	*endouc*	ZDB-GENE-060503-141
7.76E-95	1.79844025	Hb09	*tac1*	ZDB-GENE-060503-904
2.17E-158	1.62781472	HB10	*bsnb*	ZDB-GENE-120628-1
1.33E-123	1.76522582	HB10	*ppp1r1c*	ZDB-GENE-040718-273
1.10E-115	1.25526355	HB10	*snx8a*	ZDB-GENE-031202-1
2.12E-144	0.9108047	HB10	*hlfb*	ZDB-GENE-110420-3
0	5.157643	Hb11	*pyya*	ZDB-GENE-980526-71
1.69E-302	3.33837838	Hb11	*cbln1*	ZDB-GENE-040718-226
1.55E-160	2.16490785	Hb11	*sox1b*	ZDB-GENE-060322-5
2.10E-159	2.23259612	Hb11	*phc2a*	ZDB-GENE-020312-1
3.15E-142	2.00260975	Hb11	*zfhx4*	ZDB-GENE-070530-3
2.32E-111	2.13785726	Hb11	*cpne4a*	ZDB-GENE-060503-427
5.67E-125	1.70177706	Hb11	*drd4a*	ZDB-GENE-070112-996
1.06E-293	1.34716	Hb11	*sox1a*	ZDB-GENE-040718-186
6.65E-256	1.4655357	Hb12	*htr1aa*	ZDB-GENE-071203-1
8.27E-137	3.04732314	Hb12	*hbegfb*	ZDB-GENE-070820-6
3.31E-241	1.4732662	Hb12	*bmp2b*	ZDB-GENE-980526-474
4.51E-169	2.26359047	Hb13	*tubb5*	ZDB-GENE-031110-4
1.00E-87	1.1285375	Hb13	*fxyd6l*	ZDB-GENE-071205-8
2.11E-64	1.0978017	Hb13	*rpl26*	ZDB-GENE-040426-2117
9.94E-133	5.37707535	Hb14	*galn*	ZDB-GENE-111117-2
1.16E-100	2.8979173	Hb14	*smox*	ZDB-GENE-031201-3
4.01E-100	5.13196808	Hb14	*agrp*	ZDB-GENE-040817-1
1.76E-94	3.49999355	Hb14	*rrad*	ZDB-GENE-030131-5607
6.90E-64	2.59960022	Hb14	*gadd45bb*	ZDB-GENE-050223-1
1.75E-62	2.34113759	Hb14	*gad2*	ZDB-GENE-030909-9
2.50E-35	3.32589767	Hb14	*rgs4*	ZDB-GENE-030131-9839
7.05E-33	2.55014141	Hb14	*penka*	ZDB-GENE-030729-31
7.02E-30	1.703852	Hb14	*gad1b*	ZDB-GENE-030909-3
7.02E-33	1.537209	Hb14	*slc32a1*	ZDB-GENE-061201-1
2.04E-31	1.491485	Hb14	*dlx5a*	ZDB-GENE-990415-49
2.04E-21	3.28811127	Hb14	*nmu*	ZDB-GENE-041001-111
4.58E-80	3.81719398	Hb15	*kiss1*	ZDB-GENE-080128-1
2.29E-77	4.51223746	Hb15	*si:dkey-117i10.1*	ZDB-GENE-091204-371
7.39E-74	3.02185606	Hb15	*lmo3*	ZDB-GENE-050522-201
1.30E-60	3.00911414	Hb15	*csrp2*	ZDB-GENE-040426-1863
8.83E-58	2.66562076	Hb15	*wnt11r*	
6.01E-57	2.48047347	Hb15	*ak5*	ZDB-GENE-030131-8256
3.25E-51	2.32214249	Hb15	*prkcq*	ZDB-GENE-041210-195
5.56E-51	2.37258874	Hb15	*aoc1*	ZDB-GENE-061103-112
1.08E-44	2.32644535	Hb15	*cabp7*	ZDB-GENE-101112-4
2.59E-39	2.24361694	Hb15	*crip1*	ZDB-GENE-041111-1
5.43E-38	2.03810874	Hb15	*id4*	ZDB-GENE-051113-208

Table A.1 (Continued).

5.08E-36	2.19340796	Hb15	*ppp1r14ab*	ZDB-GENE-040426-1783
5.00E-34	2.13193874	Hb15	*necab1*	ZDB-GENE-050417-395
8.58E-31	2.24706434	Hb15	*alcama*	ZDB-GENE-990415-30
8.48E-30	2.13454652	Hb15	*kiss1rb*	ZDB-GENE-060526-54
0	5.6289242	Olf	*pvalb5*	ZDB-GENE-050417-336
0	5.26697673	Olf	*s100z*	ZDB-GENE-050522-69
0	4.90646722	Olf	*icn*	ZDB-GENE-030131-8599
0	4.56962686	Olf	*calb2a*	ZDB-GENE-040426-1677
0	4.14434954	Olf	*gstp1*	ZDB-GENE-020806-4

Table A.2. Curated Markers for the Adult Habenula.

Cluster	Gene	avg_diff	p_val
Ad_Hb01	*tac3a*	2.44722984	0
Ad_Hb01	*g0s2*	2.24103739	0
Ad_Hb01	*fxyd1*	2.09698509	0
Ad_Hb01	*nrgna*	1.85522745	0
Ad_Hb01	*kctd4*	1.84863675	0
Ad_Hb01	*adcyap1b*	1.7226187	0
Ad_Hb01	*tgfbi*	1.28744007	2.70E-305
Ad_Hb01	*pkdccb*	1.30468338	2.50E-235
Ad_Hb01	*nr0b1*	1.40459654	8.16E-181
Ad_Hb01	*sst1.1*	1.61857087	3.01E-121
Ad_Hb02A	*rac2*	1.62551156	8.58E-49
Ad_Hb02A	*gng8*	1.12110562	1.01E-40
Ad_Hb02A	*cyth4a*	1.12059366	4.54E-40
Ad_Hb02A	*adrb2a*	1.13151349	2.13E-39
Ad_Hb02A	*kctd1*	1.12730464	2.67E-35
Ad_Hb02A	*fam19a2*	1.15683647	4.61E-33
Ad_Hb02A	*tgfbi*	1.23037644	1.05E-30
Ad_Hb02A	*si:ch211-215d8.2*	1.19119631	4.60E-27
Ad_Hb02A	*rpz5*	1.34090933	8.16E-25
Ad_Hb02A	*pde6h*	1.54942573	1.05E-22
Ad_Hb02B	*gng2*	2.14258764	0
Ad_Hb02B	*tac3a*	1.00440902	1.22E-218
Ad_Hb02B	*neurod1*	1.39895855	9.96E-187
Ad_Hb02B	*nwd2*	1.03753216	1.39E-132
Ad_Hb02B	*kctd12.2*	0.91909188	8.75E-130
Ad_Hb02B	*syt6b*	1.03993435	1.46E-126
Ad_Hb02B	*rprma*	1.195854	2.84E-118
Ad_Hb02B	*lmo1*	1.0575783	4.62E-117
Ad_Hb02B	*bcl11ba*	0.99098964	9.46E-113
Ad_Hb02B	*drgx*	1.06212687	1.75E-87
Ad_Hb04	*cbln2b*	2.58205479	1.80E-133
Ad_Hb04	*kctd12.1*	1.0150766	6.34E-51

Table A.2 (Continued).			
Ad_Hb04	*cart2*	1.38071637	3.18E-49
Ad_Hb04	*kctd12.2*	0.93777504	5.34E-32
Ad_Hb04	*lmo1*	0.98192636	3.86E-28
Ad_Hb04	*neurod1*	1.08057742	2.98E-26
Ad_Hb04	*tenm3*	1.22363021	2.47E-24
Ad_Hb04	*kctd4*	1.11344964	2.02E-22
Ad_Hb04	*tnfrsf9a*	0.97374113	7.82E-19
Ad_Hb04	*sst1.1*	0.96422791	1.22E-12
Ad_Hb05	*trh*	3.03045398	2.79E-96
Ad_Hb05	*spa17*	3.54440756	2.76E-89
Ad_Hb05	*sst1.1*	3.1010826	5.73E-76
Ad_Hb05	*pyyb*	2.81215701	9.17E-68
Ad_Hb05	*spx*	3.23627761	4.99E-55
Ad_Hb05	*ralyl*	2.03679233	1.40E-54
Ad_Hb05	*qkia*	1.92040545	1.02E-51
Ad_Hb05	*zgc:101731*	1.78147089	5.04E-43
Ad_Hb05	*tgfbi*	1.69315831	3.74E-37
Ad_Hb05	*inaa*	1.61376345	2.23E-23
Ad_Hb06	*lrrtm1*	2.40440426	0
Ad_Hb06	*im:7152348*	2.38501208	0
Ad_Hb06	*cckb*	1.86832655	0
Ad_Hb06	*kctd12.1*	1.40287132	4.20449610901e-321
Ad_Hb06	*ralyl*	2.00521577	1.69E-239
Ad_Hb06	*pou3f1*	1.73467475	1.66E-219
Ad_Hb06	*rpz5*	1.53508378	3.81E-194
Ad_Hb06	*cygb1*	1.45158799	1.24E-166
Ad_Hb06	*foxa1*	1.5924108	6.14E-151
Ad_Hb06	*mt2*	1.36589083	1.72E-81
Ad_Hb08	*cckb*	1.41137168	1.36E-126
Ad_Hb08	*lmo1*	1.53004297	3.53E-123
Ad_Hb08	*wnt7aa*	1.38658079	8.35E-105
Ad_Hb08	*nwd2*	1.31477282	6.21E-97
Ad_Hb08	*ngb*	1.52793499	1.98E-96
Ad_Hb08	*pcp4a*	1.3575629	1.80E-93
Ad_Hb08	*nrp1a*	1.3514505	2.29E-91
Ad_Hb08	*tp53i11b*	1.39396329	7.52E-85
Ad_Hb08	*lmo4b*	1.34891295	5.99E-84
Ad_Hb08	*vps8*	1.31455898	3.29E-76
Ad_Hb09	*adcyap1a*	3.68614832	0
Ad_Hb09	*zgc:101731*	2.3421248	7.08E-297
Ad_Hb09	*pdyn*	2.15198188	9.55E-231
Ad_Hb09	*tac1*	2.89311409	8.18E-222
Ad_Hb09	*cckb*	1.84067377	7.96E-207
Ad_Hb09	*inaa*	1.84777337	1.46E-204
Ad_Hb09	*rasd1*	1.9488556	1.24E-178
Ad_Hb09	*pcp4a*	1.64467108	7.07E-152
Ad_Hb09	*igf2a*	1.72479693	6.48E-145
Ad_Hb09	*zgc:110340*	1.44381567	2.71E-80
Ad_Hb10	*spa17*	3.73080702	9.15E-79
Ad_Hb10	*pvalb7*	3.64389092	8.46E-77
Ad_Hb10	*map1b*	2.54583584	2.82E-59
Ad_Hb10	*capgb*	1.99205934	6.55E-50

Table A.2 (Continued).

Ad_Hb10	*cckb*	2.7436949	3.98E-49
Ad_Hb10	*igf2a*	1.86745052	2.70E-46
Ad_Hb10	*zgc:101731*	1.93596514	1.00E-38
Ad_Hb10	*pdyn*	1.99682615	1.68E-37
Ad_Hb10	*cxcl14*	1.85672529	1.29E-30
Ad_Hb10	*glulb*	2.18424155	3.39E-27
Ad_Hb11	*pyya*	4.70634802	7.72E-81
Ad_Hb11	*cbln1*	2.80299617	2.73E-64
Ad_Hb11	*sox1b*	2.11713167	9.05E-62
Ad_Hb11	*cpne4a*	2.59239487	7.03E-55
Ad_Hb11	*sox1a*	1.67906494	9.30E-43
Ad_Hb11	*phlda2*	1.93545892	1.33E-27
Ad_Hb11	*elavl4*	1.57365537	7.86E-24
Ad_Hb11	*cox4i2*	1.99771714	2.80E-22
Ad_Hb11	*her9*	1.94789283	2.82E-22
Ad_Hb11	*atp1b1b*	1.43059384	2.15E-18
Ad_Hb13	*stmn1b*	1.63316211	1.36E-138
Ad_Hb13	*gap43*	2.2794112	1.94E-118
Ad_Hb13	*tubb5*	2.17525365	6.95E-93
Ad_Hb13	*tuba1c*	1.10147886	5.77E-61
Ad_Hb13	*tuba1a*	1.35927837	1.53E-53
Ad_Hb13	*tmsb*	1.56879678	2.37E-46
Ad_Hb13	*marcksb*	1.1999096	8.30E-42
Ad_Hb13	*ppp1r14ba*	1.05100663	8.96E-40
Ad_Hb13	*sox11a*	1.42264698	2.50E-38
Ad_Hb13	*cnp*	1.05465716	8.43E-27
Ad_Hb14	*gad2*	2.73407918	1.44E-232
Ad_Hb14	*slc32a1*	2.35831854	2.25E-183
Ad_Hb14	*gad1b*	2.55616368	2.61E-183
Ad_Hb14	*gap43*	2.39379392	7.36E-121
Ad_Hb14	*elavl4*	2.0627057	2.45E-119
Ad_Hb14	*gata3*	1.99403415	3.06E-100
Ad_Hb14	*tal1*	1.97932646	1.61E-90
Ad_Hb14	*slc6a1a*	1.69445181	9.80E-89
Ad_Hb14	*slc6a1b*	1.8243573	1.87E-79
Ad_Hb14	*atp1b1b*	1.88404703	1.68E-77
Ad_VHb01	*cntnap2a*	1.70548938	0
Ad_VHb01	*kiss1*	1.61092335	0
Ad_VHb01	*alcama*	1.56455057	0
Ad_VHb01	*aoc1*	1.55097374	0
Ad_VHb01	*id4*	1.5314773	0
Ad_VHb01	*lmo3*	1.48507568	0
Ad_VHb01	*ak5*	1.45261531	0
Ad_VHb01	*si:dkey-117i10.1*	1.35959225	0
Ad_VHb01	*necab1*	1.33507085	0
Ad_VHb01	*ppp1r14ab*	1.29793779	0
Ad_VHb_02	*rgs5b*	2.58775792	0
Ad_VHb_02	*mcl1a*	1.51493726	1.0705299e-310
Ad_VHb_02	*egr4*	1.65745152	3.38E-252
Ad_VHb_02	*mprip*	1.92182731	5.58E-213
Ad_VHb_02	*pim1*	1.31371276	3.53E-206
Ad_VHb_02	*rergla*	1.47832147	1.57E-203

Table A.2 (Continued).			
Ad_VHb_02	*dnajb5*	1.56619012	1.14E-184
Ad_VHb_02	*dusp2*	1.46250776	4.74E-112
Ad_VHb_02	*nts*	1.49442118	2.11E-103
Ad_VHb_02	*fosb*	1.33718458	4.37E-98
Ad_VHb_03	*cd82a*	2.11749224	2.60E-184
Ad_VHb_03	*ca8*	1.33230142	2.13E-77
Ad_VHb_03	*junba*	1.29493608	5.04E-77
Ad_VHb_03	*ier2*	1.31069424	4.32E-74
Ad_VHb_03	*onecut1*	1.14566539	2.57E-73
Ad_VHb_03	*rergla*	1.20535413	4.22E-69
Ad_VHb_03	*plk2b*	1.38645754	1.00E-60
Ad_VHb_03	*rgs5b*	1.21489834	2.62E-51
Ad_VHb_03	*nr4a1*	1.15380745	1.10E-46
Ad_VHb_03	*atf3*	1.15012263	3.41E-32
Ad_VHb_04	*zgc:173443*	3.01383101	3.49E-182
Ad_VHb_04	*si:ch211-251f6.7*	2.26987055	1.42E-164
Ad_VHb_04	*mprip*	1.65923661	6.43E-25
Ad_VHb_04	*rergla*	1.3417044	1.11E-23
Ad_VHb_04	*mcl1a*	0.86309876	3.14E-23
Ad_VHb_04	*rgs5b*	1.42920437	6.00E-14
Ad_VHb_04	*cabz01063757.1*	1.02732584	8.40E-14
Ad_VHb_04	*fam13b*	0.94813222	3.45E-13
Ad_VHb_04	*egr4*	0.84789345	4.01E-12
Ad_VHb_04	*fosb*	1.06725521	7.30E-10
AdHb_16	*her4.2*	4.70836249	0
AdHb_16	*fabp7a*	4.575113	0
AdHb_16	*her4.1*	4.14555086	0
AdHb_16	*her15.2*	4.04759396	5.31E-288
AdHb_16	*her15.1*	3.86516097	1.03E-280
AdHb_16	*si:ch211-193l2.4*	4.05370179	1.91E-258
AdHb_16	*mdka*	3.50619513	8.67E-258
AdHb_16	*si:ch211-193l2.5*	3.60269108	5.78E-242
AdHb_16	*her4.4*	3.9874271	1.56E-235
AdHb_16	*si:ch211-193l2.3*	3.71713126	9.49E-229

Table A.3. Primers used to generate *in situ* probes for the *in situ* atlas.

Genes	Forward Primer	Reverse Primer	Linearize	Transcribe
tac3a	ccctgtctctgtgtcttgtctg	gcctataacccacgacgaaac	NotI	T3
tac1	acgacgtttagcttgcgttt	accgggagagatgttgtgac	NotI	T3
mglur8b	accatgttgagtgcctttg	aggatgttcgctcatatgg	NotI	T3
nrgna	aggacgaggacatcatggac	agagtatttggctcgccttg	NotI	T3
endouc	attttggatggatccctgtg	ggtggtggtcttctcagtgg	ClaI	T7
gpr139	tcctgctgctgattgtgttc	ccgttaccaaatgcaggagt	ClaI	T7
prkcq	*cagatttgacctcccacgat*	*gcttcgcaaccttttattgg*	ClaI	T7
aoc1	catttcagggagggaaaaca	ggctgtgaaaacatcacaattt	ClaI	T7
wnt7aa	ccgctggatttttcacattt	cgcgagtactggtgtgtgtt	NotI	T3
pnoca	gaagctggcgaaaggagaat	gctttgtttgacagattagatgc	ClaI	T7
adrb2a	atcgttacatcgccatcaca	aaatcaggaccagtggaacg	NotI	T3
trh	ggtgtgtgtgttcgtcttgg	ggtggcatgctgcctatatt	NotI	T3
pou3f1	accacttatccccgctctct	attggtgtcccgtaacgtgt	ClaI	T7
pyya	tctgtgtctcgggacatttg	tgttcagtgctcgtgaaagc	ClaI	T7
slc32a1	acaagcccagaatcactgct	aatccggcaggttgtatttg	ClaI	T7
lrrtm1	atatcccccaaaacctgagc	acttccccaccctccatatc	NotI	T3
foxa1	attctcccatgacgaacagc	acaggcctggaatacacacc	Cla1	T7
ppp1r1c	cagtttgccgttccactttt	aggccagaaatgcagaagaa	NotI	T3
cbln2b	gtggcatagggctctgtctg	ggagtatttccatccgccca	ClaI	T7
nptx2a	atttgtggcactgctgcttg	gctctccgtcctgataagcc	NotI	T3
fabp7a	tctctcaacatggtcgatgc	tcccacctctgaacttggac	NotI	T3
igf2a	agcaggccgaacagatcaaa	ttctgtcttcgtgccacagg	NotI	T3
murcb	gctggagaacaccgtcaaga	ctgtcaggcctgcttaagct	ClaI	T7
pcdh7b	gggctctggacagagaacag	gttgacacggatgtcacctg	ClaI	T7
wnt11r	tctcctctcaggctcagctc	tatcgcacttcttgcaggtg	ClaI	T7
penka	tgatgaactcctggtggact	tccatgaatcctccgtatctc	NotI	T3
sox1b	tctccaggagcccagactaa	gttggggtacagcgtgaagt	NotI	T3
cpne4a	attctctgggcctgttcagc	ccaaacgccgggaacatttt	NotI	T3
c1ql4b	tggttctggtggtggaggta	ttccaccgtgcacttttcct	ClaI	T7
nebl	gcgctgtggaaaaatcgtct	caacatagttggcaggcagc	NotI	T3
cd82a	ctgcgaccaagtacttcctgt	acactcttgcaaaggcccat	NotI	T3
zgc:173443	acgacttgctgaatcctgct	aacacagacaacccacagca	NotI	T3
tacr3l	acgcacggagagtggtactt	cattgggatttggtgctcftt	NotI	T3
adcyap1a	tgattacgagcagcaaaacg	ccgtggccagatacttcttc	ClaI	T7
spx	atgaaagatttaaggactcttgcgg	gtcagtacggctcgtcttcatc	HindIII	T7
tubb5	accggaacctatcacggaga	gcgtatctccgagcagttca	ClaI	T7
mprip	tccatcatctggcaggagga	cggaggtatgagagcgtgac	SpeI	T3
grm8b	tccttccggaggatttttct	tggcaggttctgtctccttt	NotI	T3
pdyn	gcaaaccgtggatggataac	atcccccgtatcgttttacc	NotI	T3
hlfb	cgcagtctgctttcttaggg	ggccatgttctctttctcca	NotI	T3
galn	atgcacaggtgtgtcggtg	ttagggttgactgatctcttctga	NotI	T3

Appendix 3: Supplementary Materials for Chapter 3

Table A.4. Allele Information for CRISPR mutants.

gpr139	
WT *gpr139*	MEHSHIFTVLTTNSSSWSPRGCPLGQFPVVYYSSLLCLGLPANILTVIILSQLVLRRQKS SYNYLLALAVADILVLLLIVFVDFLLEDFILGAPLPHSLNKAVQVLEFSSIHTSIWITVP LTIDRYIAVCHPLRYHTVSYPARTRKVILAVYAGCLISSVPYYWWPELWHGLPGASPGGR SSSAGQHVLVWVHCATVYLLPCSVFFSLNAIIVRKLRCRRSCFRLRGYSTGKTTAILLAI TSVFAVLWAPRTLMILYHLYTVQPAMPGPARLLHLVTDVANMLALLNTGVNFFLYCFISK RFRRMAGTVLKAFFRCRKQPPPFYASHNFSITSSPWISPANSHCIKMLVYQYDKNGKPVC ISS
gpr139 (226bp del)	MEHSHIFTVLTTNSSSWSPRGCPLGQFPVVYYSSLLCLGLPANILTVIILSQLVLRRQKS SYNYLLALAVADILVLPLGA-Stop
gRNA Target Sites	gpr139_1: *TAATACGACTCACTATA*ggACAACTACCTACTGGCCCGTTTTAGAGCTATGCTGGAAAC gpr139_2: *TAATACGACTCACTATA*ggCATCTTGGCGGTGTACGCGTTTTAGAGCTATGCTGGAAAC gpr139_3: *TAATACGACTCACTATA*ggAGCGTAGTGGATGGCACAGTTTTAGAGCTATGCTGGAAAC gpr139_4: *TAATACGACTCACTATA*ggTGCTGCTGATTGTGTTCGGTTTTAGAGCTATGCTGGAAAC
Genotyping Primers	Forward: 5’- GTGATCATCCTTTCTCAGCT-3’ Gpr139_reverse_wt: 5’- AAGTCCACGAACACAATCAG-3’ Gpr139_reverse_mut: 5’-GACCCACACTAGCACATG-3’
Genotyping size	Mutant: 227 Wild Type: 117
aoc1	
WT *aoc1*	MWFSWMLLLATLASSSASSRTREWAHHGAMMFADLTPQEMYAVRDYLYSCSELGLTSARG TSLKKNSILLMELHVPRKHEALRALDKGQAKPSRQARVVVQFGNQAVPNVTEYIVGPLPF PKTYHLKTFKNNKNIRFESRPISAVEYEHLSGVLDKVGSKVNKILQESTGFTYGNCTKRC LTFSDIAPRGLTSGERRTWIMLQKFVEGYFIHPVGFEVLVNHKDLDHEKWTVEKVWYNGQ YFDSLDEFVEKYEKGTVDKIKLPEHDEEDLFSTYIPRGDMNTRTNIHGAKLVEPQGRRFQ VDGNFVEYAGWSFAYRVRSSAGLQIFDLRYNGERIAYEIALQEAIAFYSGDTPAAMQTKY IDAGWAMGTSDYELSPGIDCPEIAHFVDLYHYYDTDKPVRYRNALCIFEMTTGIPLRRHF NSNFQGGYNFYGGLENHVLVIRTTSTVYNYDYIWDFVFYQNGVMESRVSATGYIHATFFT ENGLNYGTRVYNYVLGNLHTHLIHYKVDLDISGRENSFESIDLKYVNFTNPWSPGHTIMQ SKLHRTQYETERSAAFRFGKKFPKYLHFYNPNQLNKWGHKKGYRIQYNSHANSVLPRGWR EENGIPWSRYPLAVTRHKDSEVTSSSIYTQNDPWEPLVSFEEFVRNNENIVNQDLVAWVT VGFLHIPHSEDVPNTATPGNSVGFFLRPFNFFNEDPSLASRSTVIVRPDEKGQPKVQRWT PEVVGHCVSDKPFFYNGTYAGV
aoc1 (28bpD, 92bpD)	MWFSWMLLLATLASSSASSRTREWAHHGAMMFADLTPQEMYAVRDYLYSCSELGLTSARG TSLKMRCQMSLNTSLAHFHSRRLTIStop

gRNA Target Sites	CCTCCTCATGGAGCTGCATGTTC TTCATCACGAACGCGGGAATGGG GGTGCAGTTCGGCAACCAAGCGG CACTCCGCAAGAAATGTACGCGG
Genotyping Primers	Forward: **tctttccctacacgac**GCTCTTCCGATCTCTGCTGGCAACTCTGGCTTCATC Reverse: **tggagttcagacgtgt**GCTCTTCCGATCTGTGGGCCAACGATGTATTCAGTGAC
Genotyping size	WT: 389bp Mutant: 269bp
ntsr1	
WT ntsr1	
ntsr1 (D)	
gRNA target sites	ntsr1_1: *TAATACGACTCACTATA*ggCCGTGGTCATTGCTTTTGGTTTTAGAGCTATGCTGGAAAC ntsr1_2: *TAATACGACTCACTATA*ggTGGCTGCCCTACCATGCCGTTTTAGAGCTATGCTGGAAAC ntsr1_3: *TAATACGACTCACTATA*ggTACTGCTATGTCACTGAAGTTTTAGAGCTATGCTGGAAAC ntsr1_4: *TAATACGACTCACTATA*ggATTGCTTTTGTGGTGTGCGTTTTAGAGCTATGCTGGAAAC
Genotyping primers	Forward: ATGTGATACTGCTGCTTATTATGACTCAG Reverse: CCTTTGATCTTTATTGATATTTCAATAAGC
nts1	
WT nts	MCRNCGTTNNMQMQLTSFLLLFLLCNGLCSDIDQGKRAIEDEVLRSLLTSKVKASRHIAP LWQLPLQDVCRMVNGLGDSWLEAWANEEAAEDTEVHADYEQRVSGTLLQMLEEMHDIQNL CRVLQPRELQDEQEYLELEQNSDSPLKRKSPYILKRQLRTNKSRRPYILKRSVIY
nts(D)	
gRNA targets	nts_1: *TAATACGACTCACTATA*ggAGAGGAAGCTGCCACAGGGTTTTAGAGCTATGCTGGAAAC nts_2: *TAATACGACTCACTATA*ggGGTTGCAGGACCCTGCAGGTTTTAGAGCTATGCTGGAAAC nts_3: *TAATACGACTCACTATA*ggTATGAGCAGAGGGTCTCGGTTTTAGAGCTATGCTGGAAAC nts_4: *TAATACGACTCACTATA*ggGGCTCGGTGACTCGTGGCGTTTTAGAGCTATGCTGGAAAC
Genotyping Primers	Forward: **tctttccctacacgac**GCTCTTCCGATCTCATCTGTGCAGGTAAAGGCGAGC Reverse: **tggagttcagacgtgt**GCTCTTCCGATCTATTTACAGTCTCAAATAAATGTCTCACC

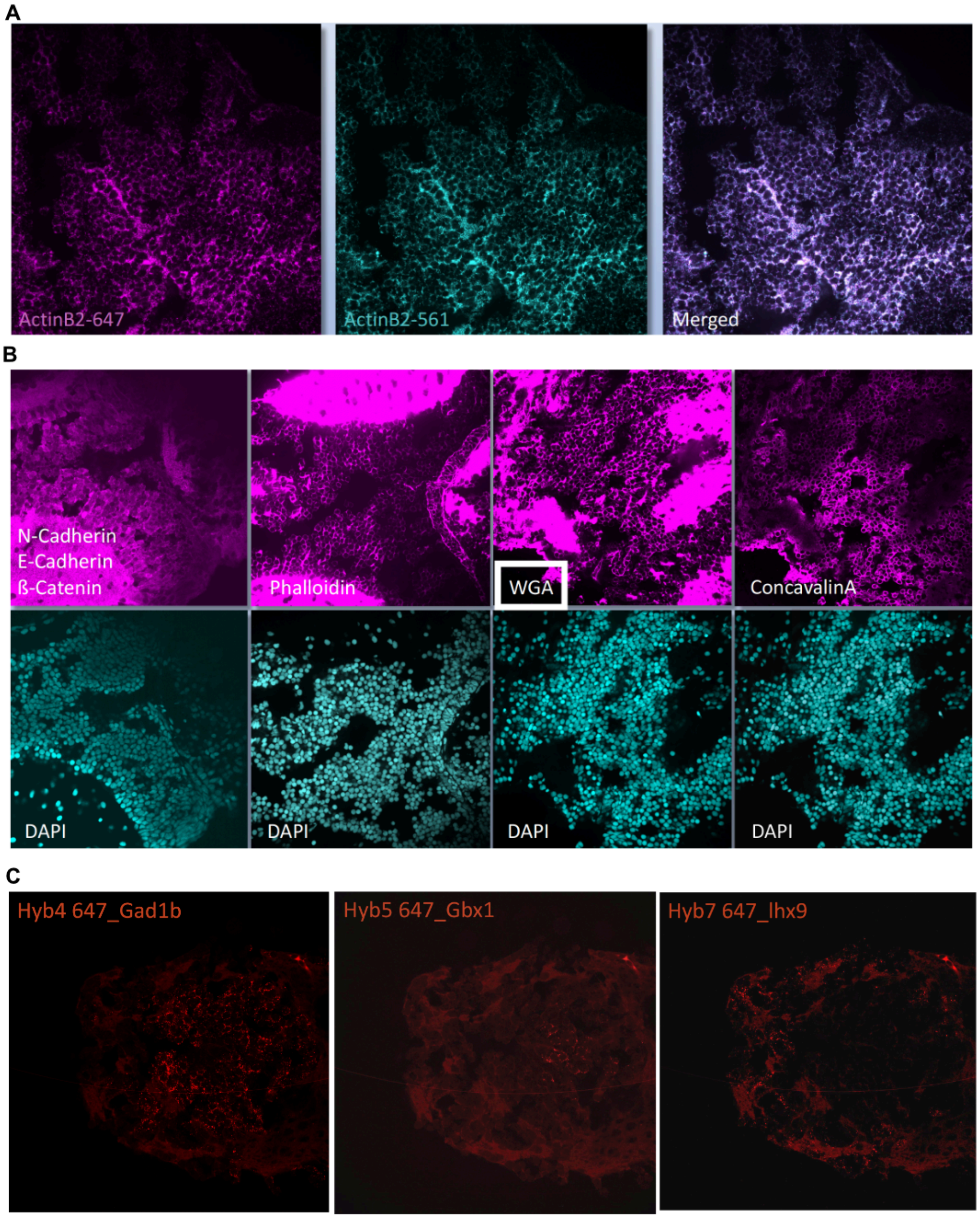


Figure A.6. Preliminary Seq-FISH Optimization Data.

(A) Colocalization of signal across two rounds of hybridization to assess RNA integrity.

(B) Optimization of a cell segmentation marker for segmentation of individual cells during automated seq-FISH analysis. WGA worked best for automated cell segmentation

(C) Three example ROIs from sequential rounds of hybridization of probes for three different genes indicate that the signal amplified is consistent with the expression patterns of forebrain marker genes.

Appendix 4: Supplementary Materials for Chapter 4

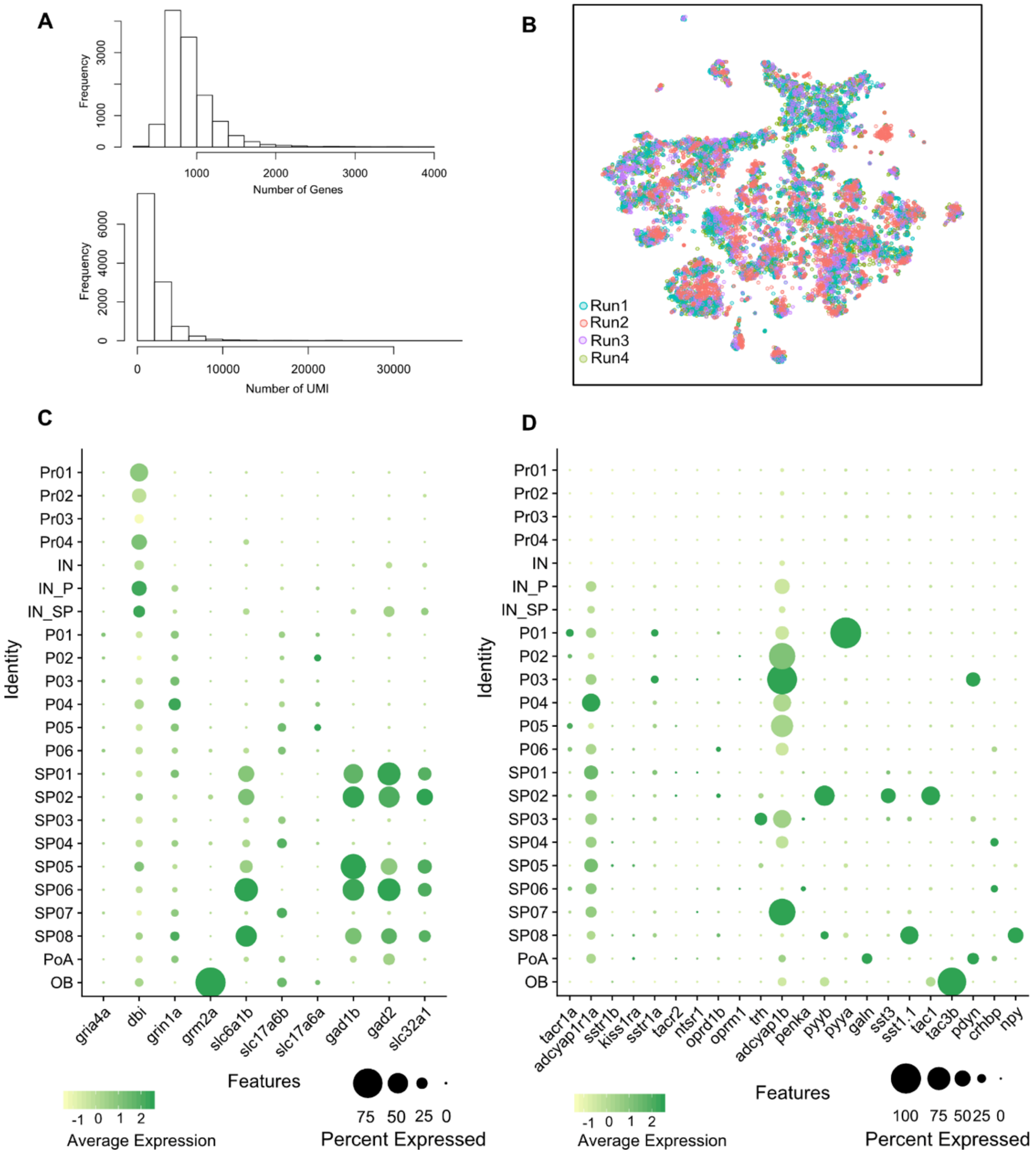


Figure A.7. Quality Control Metrics and Expression of Neurotransmitter and Neuropeptidergic Genes Among the Telencephalon Clusters.

(A) Distribution of number of genes(top) and number of UMIs(bottom) detected across the filtered larval dataset

(B) Composition of each cluster by experimental batch post batch correction and filtering

Figure A.7 (Continued). (C) Gene Expression profiles of selected neurotransmitters and neurotransmitters among the telencephalon clusters.

(D) Gene Expression profiles of selected neuropeptides and neuropeptide receptors among the telencephalon clusters.

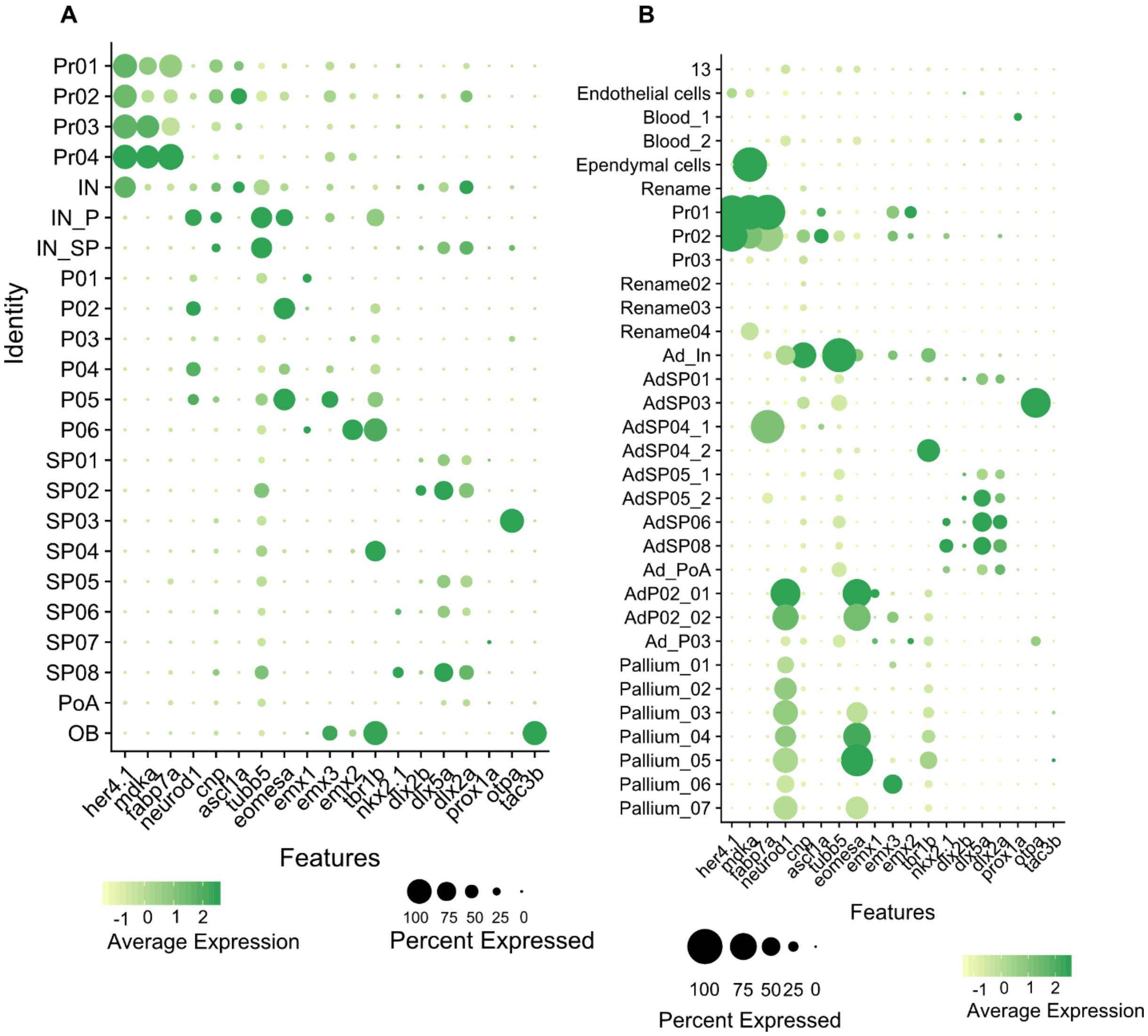


Figure A.8. Gene Expression profiles of previously known molecular markers of in the A) larval and B) adult telencephalon.

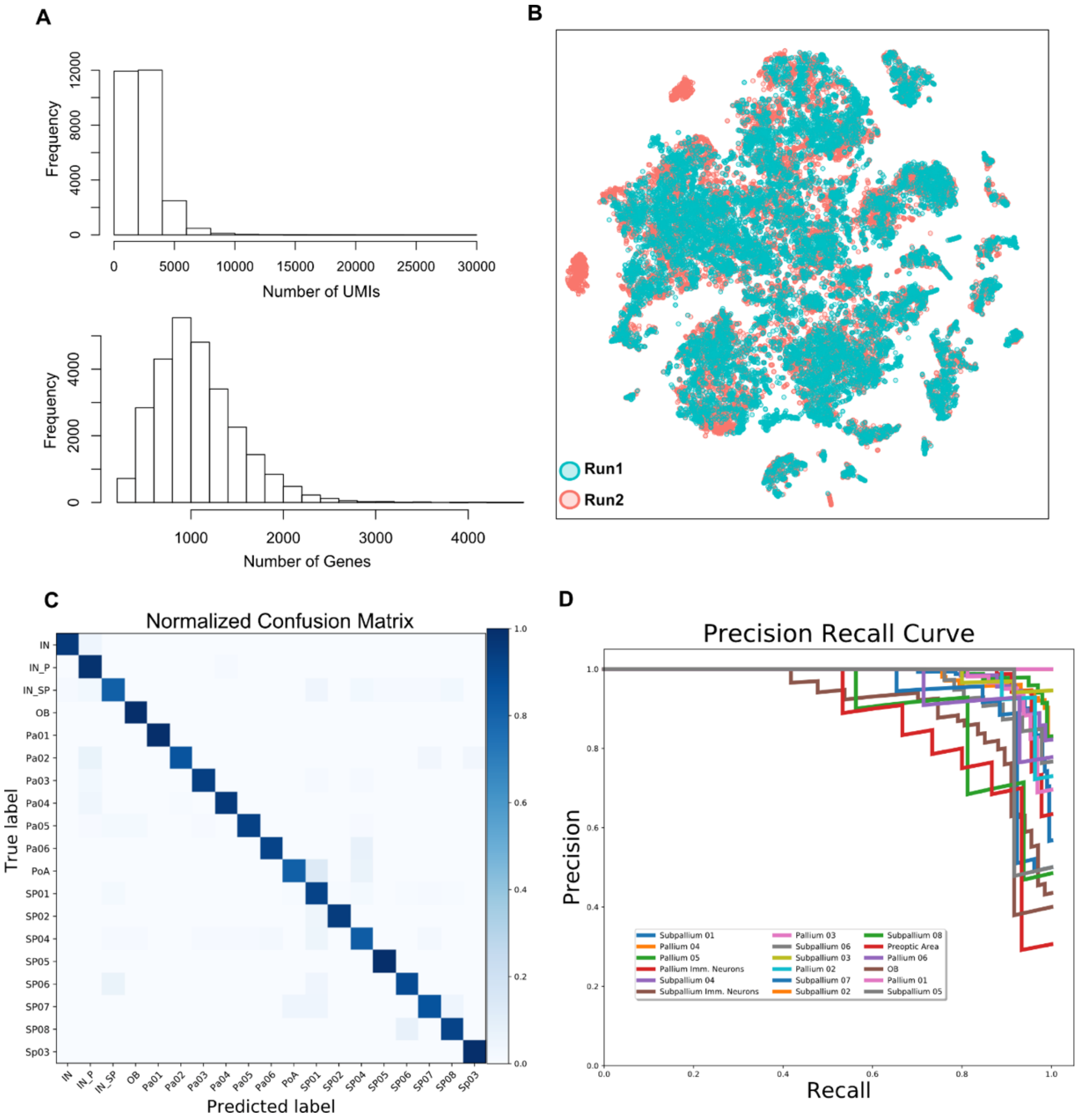


Figure A.9. Quality Control Metrics for the Adult Dataset and Performance of Classifier Trained on the Larval Data.

(A) Distribution of number of genes(top) and number of UMIs(bottom) detected across the filtered adult dataset

(B) Composition of each cluster by experimental batch post batch correction and filtering

(C) Performance of Gradient Boosted Trees (GBT) model trained on the larval dataset with the graph clustering labels shown in Figure 1A. A training set was formed by choosing 90% of the cells from the entire dataset, with proportional representation from each cluster. The trained GBT model was then used to classify each cell in the remaining 10% of the data (test set) into learned cluster labels. The result of the classification of the test set is shown as a confusion matrix.

(D) Precision-recall curves for each of the class labels in the larval dataset demonstrating high sores for both precision and recall suggesting that the classifier is returning accurate results as well as a majority of all positive results.

Table A.5: Primers used to generate *in situ* probes for the *in situ* atlas (f: forward; r: reverse)

Gene Name	Primers
bhlhe22_f	GTTTCGACCGACTGCGGGTA
bhlhe22_r	AATCCCACGCTCACCAGCAC
c1ql3b_chr2_f	GGTGCTGGTGCTGGTGATCC
c1ql3b_chr2_r	GGTGCTGGTGCTGGTGATCC
cbln1_f	CCATGCATACGCACAGAACGAG
cbln1_r	CAATGGCGTTTGAGACTCACGGT
cnp_f	GTGTTTTGTTTGTCAGAAGCAGTGT
cnp_r	GTCTGAACCGCCTCCACACC
c1ql3b_chr3_f	ATGCTGGTGCTGGTGCTGG
c1ql3b_chr3_r	CCGTGGGCTTTGCCACCAT
crabp1a_f	ATGCCTCCCAACTTCGCCG
crapp1a_r	TGTGCCAGAAATGGACAAAGAGA
crhbp_f	TGTGCTTCCTCCTGTTGAGCG
crhbp_r	GCTGAAGCTCTTGGTGGCCC
cxcl12b_f	CAAAGTAGTAGCGCTGGTGGCG
cxcl12b_r	CCACAAAGCAAGCATTAGCATGAGC
ebf3a_f	TGCAGTCGCTGCTGTGACAA
ebf3a_r	GCCATGGTGGGACTGGAAGG
efna1b_f	CGCATGGCGAGATCGCTTCC
efna1b_r	ATCTAGCCTAGTGTTCACCCGTGTT
eomesa_f	CAACAGACCGCTGTGGCTCA
eomesa_r	CCGGCTGCCATAGATGCGAA
ets2_f	CGCATCTGCGGCATCTCTGA
ets2_r	CGCTTCCCCGAGGTCTTGTG
fezf1_f	CCAAGCTCCCGTGGGGAAAG
fezf1_r	GTGAGTGGGGTCCAGGGGAA
gabrr2b_f	GTGCCAGTGGGTGTGGATGT
gabrr2b_r	CAGCCCTCTGAGCGTCCTTT
gad1b_f	ATTAATGCGGGAAAGTGGATCTGTT
gad1b_r	GACTTGGTGACGGCGTGATTG
gbx1_f	TGCCGTACCGACCGCTTATG
gbx1_r	TGTGGCCCCTGACCAGTGAA
isg15_f	TGAGCGGTGATGCTACCGTTG
isg15_r	CGTGTGAACATCACGGCATTGAAA
lhx6_f	AGAAAGACGACACGCGCTCC
lhx6_r	CGGCTAAGAGGAAGCTGCGG
lhx8a_f	CGTCGTCGTTGTCTTTGTCGC
lhx8a_r	ATGGCTGATGGGCAGCTGTG
lhx9_f	ATGGAAGTTGTGGGGTGCAAGG
lhx9_r	CCTGTCTTTTGGGCGAGTTGTTT
mdka_f	ATGCGGGGCCTGTTTTCCAC
mdka_r	GGGAAACAGCAGAGGAGTGGC
mmped2a_f	CGGTATGCCGACACAAGGCAA
mmped2a_r	GTCCGTCATCGGAAACTTCCACTC
nfia_f	TCCGCTATGCCTAACACAGGA
nfia_r	TGGCTCCACTTCATGCCAGC
nr2f2_f	TGGCCAGTTCACTTGCGAGG

Table A.5 (Continued).

nr2f2_r	AACTGCTCCCCGACAGCAAC
nrgnb_really_spa17_f	GAGCAGCCAGAGGACATCGC
nrgnb_really_spa17_r	CCTCACCTGTACGAGTGTCGC
otpa_f	CGAGCTTCTGGGTCACAGGG
otpa_r	TTAGGTGAAGCTCATGGACACTG
penkb_f	AGACTTTACACGCACCGGACA
penkb_r	ATGAATGCCTCATGCAAACCAGAAT
ppp1r14aa_f	ACACATACCGGTCATTCCCCTGA
ppp1r14aa_r	TGCATGTTTCCGTGATGATGGG
prdm8b_f	CAAGGACGCCAGTGACCAGG
prdm8b_r	TGGCACACCAGTTCTGAGCG
prdm12b_f	ACAGTCTTTGGCCCTGTCCG
prdm12b_r	GCCTGAAGCCCCACTACAGC
prkcda_f	GAAGGCAAACGGGCATGCAG
prkcda_r	TCCCAGGCAAGAGCAAGCAC
pvalb7_f	AGCCCTTGATCAATTCAAAGCTGC
pvalb7_r	GTGTTCGGTGGCTCTATCACA
pyyb_f	ATGGCGAGTGCACTGAGATCC
pyyb_r	TGAGGGTTTGGCTCATGCTGG
rprma_f	ACCAAACGGACAGTGGCATCT
rprma_r	ACAATGCTTAAGTTTACCAGAGACC
rprml_f	TGCTGAATCGCAGTCAGGAGC
rprml_r	ACCAAATTGCTCATGCGCTTTCC
scrt1b_f	CACCCCGGCCATTTATGATGGA
scrt1b_r	CGCCAGTCATAGTCGAGTGCTG
six3b_f	CGGATCGCCCTCTGCTTCTG
six3b_r	GGGGATTCTGCCGAGCTGTG
synpr_f	CTGGGCAAATGCGGGTTAGTG
synpr_r	AATGGATTTCCCATGTGAGCGAGTC
tmsb_f	TGGCCGACAAACCCAACATGA
tmsb_r	TGCTGGTGTGTAGCACCAAA
trh_f	GGCGGTGTGTGTGTTCGTCT
trh_r	TGGGAGGTATGAAATCCGACAGC
uncx_f	CTGCACGGCTTCTGTGGTGA
uncx_r	CGCGCCGCTGTTGTTGTTTA
vim_f	AAGCTTGAGAGTGCGCTCCG
vim_r	CCTTCATGGACTCTCGCAGGC
zbtb18_f	CAGGAGCCAGGGAGCATGTG
zbtb18_r	AGTGCCCAGTCTCTGACCGT
zic2a_f	CTGCATTCACGAGCCAAGCG
zic2a_r	GTCGGATGCAGGTGGGGATG
tac3b_withutr_f	AGCGCCCTCGACTACTCCTT
tac3b_withutr_r	TGCACAAATTATGGCCAACATGC
grm2a_newlong_f	GCCCGGAGATGTTTGCTGCT
grm2a_newlong_r	ATGTTGGAGGCCGAGGCTTG
uts1_withutr_f	GCTCTTCACCTTCCCGCTCC
uts1_withutr_r	GCGCGTCCCTGAAGGTGTTT

Appendix 5: Supplementary Materials for Chapter 5

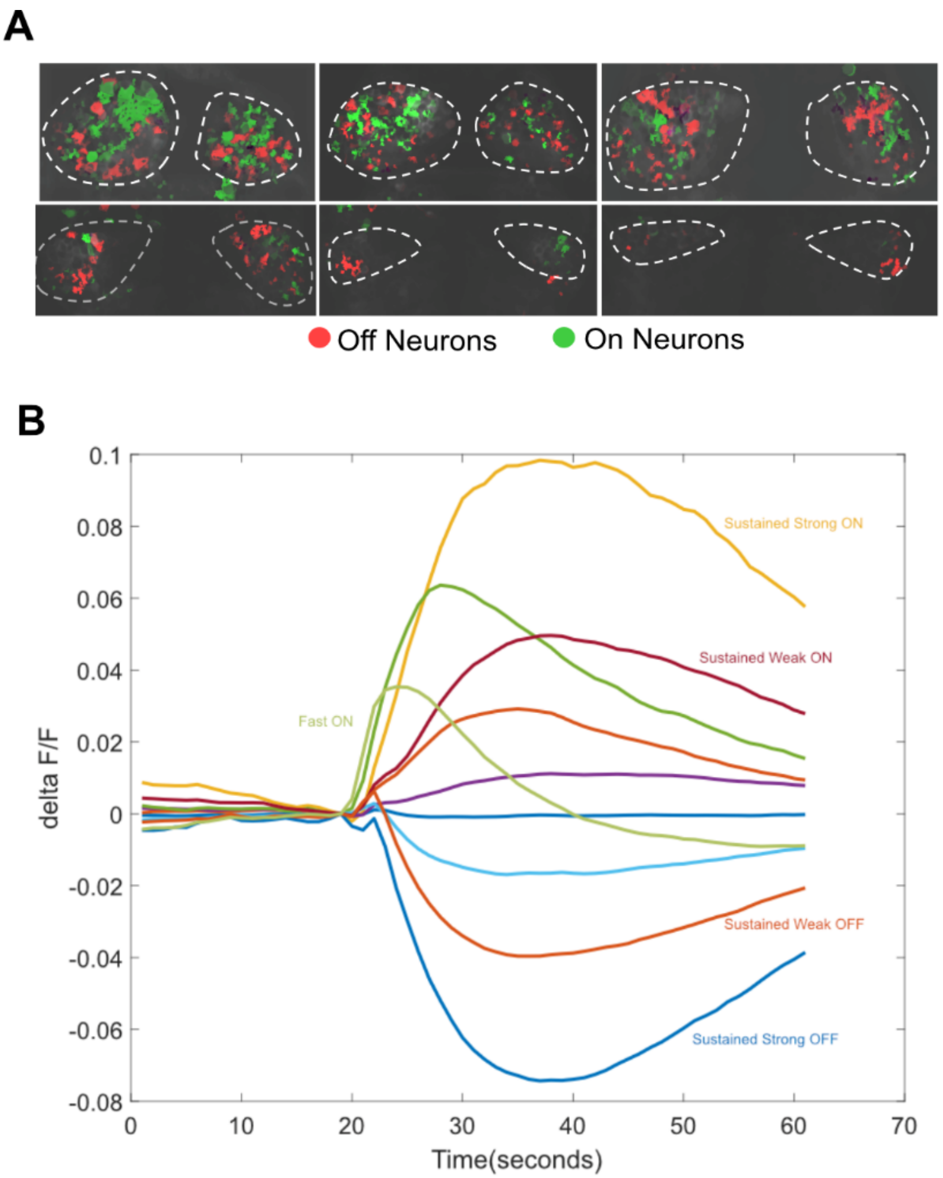


Figure A.10. Neuronal Dynamics within the Habenula in response to Noxious Electric Shocks.

(A) Distribution of shock ON and OFF cells within the habenula

(B) Different Types of response dynamics found among the on and off cells.