Comparative analysis of the molecular, spatial, and functional domains of vertebrate habenula

by

Yağnur Işık Çiftci Çobanoğlu

A Dissertation Submitted to the
Graduate School of Health Sciences
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Neuroscience

KOÇ
ÜNİVERSİTESİ

June 4, 2024

Comparative analysis of the molecular, spatial, and functional domains of vertebrate habenula

Koç University

Graduate School of Health Sciences

This is to certify that I have examined this copy of a master's thesis by

Yağnur Işık Çiftci Çobanoğlu

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Dr. Emre Yakşi (Advisor)

_____

Prof. Dr. Stephen W. Wilson

_____

Assoc. Prof. Atay Vural

Date: _____

ABSTRACT

Comparative analysis of the molecular, spatial, and functional domains of vertebrate habenula

Yağnur Işık Çiftci Çobanoğlu

Master of Science in Neuroscience

June 4, 2024

The habenula (Hb) is a diencephalic brain region implicated in adaptive behaviors and neuropsychiatric disorders. Two main Hb subdivisions, the lateral habenula and medial habenula, have been shown to play different roles in diverse behaviors, ranging from classical and operant learning to the prediction of outcomes, fear, addiction, and social interactions. However, our understanding of cytoarchitecture and topography of the vertebrate Hb, as well as how different Hb microcircuits are organized across species, remains limited. In this study, we utilized a combination of single-cell and spatial transcriptomic approaches across multiple species to investigate how habenular subdivisions are organized and how this organization compares across species in relation to Hb function. First, we demonstrated that habenular cell types are relatively conserved across species from fish to mice, with some exceptions. Subsequently, employing spatial transcriptomics at subcellular resolution, we identified numerous topographically organized subregions within the adult zebrafish and mouse Hb. Many of these subregions were found to be shared between both species. We also observed that while zebrafish Hb is known to be highly asymmetric across hemispheres, several habenular subdomains remain symmetric, resembling those of mouse Hb. Similarly, though the mouse habenula is traditionally considered highly symmetrical across hemispheres, we observed a few mouse Hb subregions with prominent asymmetries. These results suggest that the asymmetry of the vertebrate habenula exists on a continuum across species. Moreover, vertebrates share many cytoarchitectural similarities that are topographically organized. Our ongoing investigation aims to elucidate how the molecular topography of the zebrafish Hb relates to its functional architecture.

ÖZETÇE

Omurgalı habenulasının farklı moleküler, uzamsal ve fonksiyonel bölgelerinin belirlenmesi ve türler arası karşılaştırılması

Yağnur Işık Çiftci Çobanoğlu

Nörobilim, Yüksek Lisans

4 Haziran, 2024

Habenula (Hb), adaptif davranışlar ve nöropsikiyatrik bozukluklarla ilişkili olan diensefalik bir beyin bölgesidir. İki ana Hb alt bölümü, lateral habenula ve medial habenula, klasik ve edimsel öğrenmeden sonuçların öngörülmesi, korku, bağımlılık ve sosyal etkileşimlere kadar çeşitli davranışlarda farklı roller oynamaktadır. Bununla birlikte, omurgalı habenulasının hücre yapısı ve topografyası ile farklı habenula mikro devrelerinin türler arasında nasıl organize edildiği konusundaki anlayışımız sınırlıdır. Bu çalışmada, habenulanın alt bölümlerinin nasıl organize edildiğini ve bu organizasyonun habenula fonksiyonu ile ilgili olarak türler arasında nasıl karşılaştırıldığını araştırmak için tek hücre RNA dizileme (scRNA-seq) ve mekansal transkriptomik yaklaşımlarını birleştirerek farklı omurgalı türlerde araştırma yaptık. İlk olarak, habenular hücre tiplerinin balıklardan farelere kadar türler arasında , bazı istisnalar dışında, nispeten korunduğunu gösterdik. Ardından, hücre altı çözünürlükte mekansal transkriptomikleri kullanarak, yetişkin zebra balığı ve fare habenulası içinde çok sayıda topografik olarak organize edilmiş alt bölgeyi tanımladık. Bu alt bölgelerin birçoğunun her iki tür arasında paylaşıldığını bulduk. Ayrıca, zebra balığı habenulasının yarıküreler arasında oldukça asimetrik olduğu bilinirken, bazı habenular alt bölgelerin fare habenulası gibi simetrik kaldığını gözlemledik. Benzer şekilde, fare habenulası geleneksel olarak yarıküreler arasında oldukça simetrik kabul edilse de, birkaç fare habenula alt bölgesinde belirgin asimetriler gözlemledik. Bu sonuçlar, omurgalı habenulasının asimetrisinin türler arasında bir devamlılık içinde olduğunu göstermektedir. Ayrıca, omurgalılar, topografik olarak organize edilmiş birçok hücre yapısı benzerliği paylaşmaktadır. Devam eden araştırmamız, zebrafish Hb'sinin moleküler topografyasının fonksiyonel mimarisiyle nasıl ilişkili olduğunu aydınlatmayı amaçlamaktadır.

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest appreciation to my advisor, Emre Yakşi, for his amazing guidance, feedback, and patience. Thanks to him, I am fascinated with science even more and eager to continue in my academic journey. Countless thanks to Bjørn André Bredesen-Aa for his countless hours teaching and guiding me. Also, major thanks to Nathalie Jurisch-Yakşi for her support and valuable feedback.

I would like to thank Koç University and KUTTAM for accepting me into the amazing Neuroscience Graduate Program and supporting me financially.

I could not have undertaken this journey without the endless support of my friends and family, especially my parents and spouse. Thank you all for being there for me whenever I needed it, even when I did not realize it.

Huge thanks to the lab members of Yakşi lab members, with a special thanks to Berkay, who we worked with and supported each other through this project.

Table of Contents

LIST OF FIGURES

ABBREVIATIONS

AP	Anterior Posterior
AZBA	Adult Zebrafish Brain Atlas
AZTEC	Atlas of Zebrafish Transcriptomic Encephalic Cytoarchitecture
CCA	Canonical Correlation Analysis
CCF	Common Coordinate Framework
cDNA	Complementary DNA
DHB	Dorsal Habenula
DPBS	Dulbecco's Phosphate Buffered Saline
EggNOG	Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups
EPI	Epithalamus
GABA	Gamma-aminobutyric Acid
GFP	Green Fluorescent Protein
HB	Habenula
HCR	Hybridization Chain Reaction
IDE	Integrated Development Environment
IPN	Interpeduncular Nucleus
IVT	In Vitro Transcription
KMR	K Mutual Ratio
KNN	K-Nearest Neighbors
LH	Lateral Habenula
LHB	Lateral Habenula
LI	Lateralization Index
LM	Lateral Medial
MeOH	Methanol
MERFISH	Multiplexed Error Robust Fluorescence in situ Hybridization
MH	Medial Habenula
MHB	Medial Habenula
mRNA	Messenger RNA

PBS	Phosphate Buffered Saline
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PFC	Prefrontal Cortex
RMT	Rostromedial Tegmental Nucleus
RO	Reverse Osmosis
scRNA-seq	Single-cell RNA Sequencing
SNN	Shared Nearest Neighbor
SSCT	Sodium Chloride–Sodium Citrate with Tween 20
TH	Thalamus
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifiers
VD	Ventral Dorsal
VHB	Ventral Habenula
VTA	Ventral Tegmental Area

Chapter 1:

INTRODUCTION

1.1 Habenula as a nucleus

The habenula is a conserved diencephalic nucleus that is important for neurophysiology and behavior in vertebrates. This region is named after the word "habena" which means "little rein" because of its shape. In addition, it is conserved throughout evolution (Figure 1.1) (Bianco & Wilson, 2009; Fore et al., 2018; Namboodiri et al., 2016; Pandey et al., 2018). Habenula is part of the epithalamus and it is surrounded by the thalamus. It is positioned near the midline and adjacent to the third ventricle (Namboodiri et al., 2016).

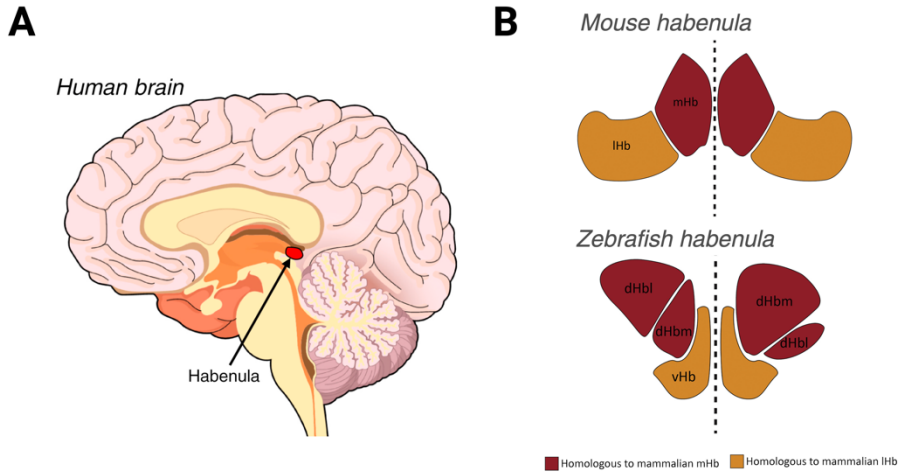

Figure 1.1: Comparative anatomy of the habenula. (A) Illustration of the human brain highlighting the location of the habenula (Namboodiri et al., 2016). (B) (Top) Anatomical illustration of adult mouse habenula, divided into its subregions: medial habenula (mHb) and lateral habenula (lHb). (Bottom) Anatomical illustration of adult zebrafish habenula, divided into its subregions: the dorsal habenula lateral (dHbl), dorsal habenula medial (dHbm), and ventral habenula (vHb). The red color indicates regions homologous to the mammalian medial habenula (mHb), and the orange color indicates regions homologous to the mammalian lateral habenula (lHb). (Schmidt, 2017), (Created with BioRender).

Previous studies have shown its significant role in behavior, operant learning, and neurophysiology in vertebrates (Fabrizio Palumbo, 2020; Hashikawa et al., 2020; Michel et al., 2024; Okamoto, 2012). Research indicates that the habenula influences a variety of behaviors, including sleep, reward learning, fear, and pain processing (Amo, 2010; Cohen

& Melzack, 1993; Michel et al., 2024; Okamoto, 2012). Furthermore, the role of the habenula is shown by previous studies in coordinating various neural activities, assessing outcomes, and choosing the optimal behaviors in reaction to changes (Fabrizio Palumbo, 2020). Additionally, abnormalities in the habenula's function are associated with neurological disorders such as addiction and depression (Proulx et al., 2014).

1.2 Molecularly defined cell types of habenula

In mammals, the habenula is divided into the medial (mHb) and lateral habenula (lHb) which have diverse afferent inputs and efferent connections (deCarvalho et al., 2014; Namboodiri et al., 2016). Hence, the neuroanatomical connectivity of mHb and lHb differs significantly. The medial habenula mainly receives signals from the medial and lateral septal nuclei and sends its output primarily to the interpeduncular nucleus in the midbrain (Bianco & Wilson, 2009; Contestabile & Flumerfelt, 1981). In contrast, the lateral habenula links various structures like the septum, hypothalamus, basal forebrain, globus pallidus, and prefrontal cortex with the dopaminergic, serotonergic, and noradrenergic systems (Bartoszek, 2021; Michel et al., 2024; Namboodiri et al., 2016). Furthermore, lateral and medial parts of the lateral habenula have different projections. Neurons that project to dopaminergic neurons are typically located in the lateral part of the lateral habenula, whereas those that project to the raphe nuclei, which contain serotonergic neurons, are usually found in the medial part of the lateral habenula (Namboodiri et al., 2016). In addition, gene expressions in mHb and lHb differ significantly. The neurons in the lateral habenula are found to be primarily glutamatergic. The medial habenula also mainly contains glutamatergic neurons, but the ventral medial habenula neurons also co-release acetylcholine. Furthermore, neurons in the dorsal medial habenula co-release the neuropeptide substance P, which is confirmed by *tac1* expression (Namboodiri et al., 2016).

Zebrafish habenula divided into the dorsal (dHb) and ventral habenula (vHb), which are found to be the homologous of the mammalian medial and lateral habenula, respectively (Amo, 2010; Okamoto & Aizawa, 2013). The zebrafish dorsal and ventral habenula exhibit molecular and anatomical differences and are involved in distinct behavioral aspects which will be discussed in further sections (Fore et al., 2018).

In the dorsal habenula of zebrafish, discrete subnuclei are found to differ in size and organization among the left and right sides of the brain (deCarvalho et al., 2014). Dorsal habenula further divides into dorsolateral domain (dHbl) and dorsomedial domain (dHbm), and anatomically there is an inherent left-right asymmetry between these domains (Masakazu Agetsuma, 2010; Ulrike Hüsken, 2014).

In a recent study, researchers analyzed the zebrafish habenula in larval and adult stages. They found that the neuronal cell types are stable between adult and larval stages, and these cell types are anatomically organized (Pandey et al., 2018). Furthermore, they identified thirteen new neuronal types in the zebrafish habenula and revealed dozens of new gene markers specific to these cell types (Pandey et al., 2018). Another study focused on the mouse habenula, specifically analyzing cell type profiling and conducting in situ hybridization experiments (Hashikawa et al., 2020). Researchers identified six distinct cell types in both the mHb and lHb domains. They found that these cell groups can be defined by several gene markers rather than singular gene expression (Hashikawa et al., 2020). Additionally, they discovered that genes expressed in the mHb are largely associated with membrane conductance and cholinergic transmission (Hashikawa et al., 2020). Within the mHb, they observed superior, inferior, central, and lateral subdomains. In the lHb, they identified anterior, parvocellular, superior, central, magnocellular, and lateral subdivisions (Hashikawa et al., 2020).

1.3 Neuro-connectivity and Function of the Habenula

The habenula connects forebrain regions like the prefrontal cortex (PFC) (Warden, 2012), septohippocampal region (Okamoto, 2012), and basal ganglia to the downstream monoaminergic nuclei, including the ventral tegmental area/rostromedial tegmental nucleus (VTA/RMTg) and the raphe nuclei, as well as the interpeduncular nucleus (IPN) (Amo et al., 2014; Fore et al., 2018; Okamoto, 2012). Furthermore, it receives input from many brain regions and affects behaviors like sleep, pain processing, reward learning, and fear (Amo, 2010; Cohen & Melzack, 1993; Michel et al., 2024; Okamoto, 2012; Pandey et al., 2018). Hence, it serves as a crucial neuroanatomical center, connecting and regulating various brain

regions that are essential for different motivational states and cognitive processes (Namboodiri et al., 2016).

Habenula dysfunction is linked to neurological disorders such as depression, schizophrenia, and addiction (Hikosaka, 2010; Lecca et al., 2014). Specifically, patients with major depressive disorder have been found to have changes in habenular volume (Schmidt, 2017) and activity (Morris, 1999). In a passive conditioning task, where a neutral cue is paired with a negative stimulus (like an electric shock), healthy individuals show increased habenular activity in response to the cue. However, depressive patients exhibit decreased habenular activity as the cue-stimulus association becomes stronger (Fore et al., 2018; Lawson et al., 2017). Additionally, deep brain stimulation of the habenula has been associated with reduced depression symptoms in a human patient (Sartorius et al., 2010), suggesting new therapeutic possibilities for treating depression by targeting habenular activity (Fore et al., 2018).

The medial habenula (mHb) is believed to play a crucial role in anxiety, fear, depression, and nicotine addiction (Shih et al., 2014; Viswanath et al., 2013; Yamaguchi et al., 2013). Furthermore, it is shown that circadian rhythm can regulate the mHb activity (Sakhi et al., 2014).

Numerous studies have explored the role of the lateral habenula (lHb) in brain function and animal behavior, particularly its influence on dopaminergic and serotonergic nuclei, which helps shape both innate and learned behaviors (Fore et al., 2018). Furthermore, the lateral habenula influences brain systems that control mood, motivation, thinking, and social behavior. It plays a key role in helping midbrain dopamine neurons signal reward prediction error (Namboodiri et al., 2016).

The lateral habenula can reduce dopamine activity in the midbrain by sending signals to specific areas that inhibit these neurons. It also sends direct signals to some midbrain dopamine neurons to regulate their activity (Namboodiri et al., 2016). Moreover, the lateral habenula is the only known brain structure where nearly all incoming signals co-release both glutamate and Gamma-aminobutyric acid (GABA) (Namboodiri et al., 2016). The neuroanatomical connections of the lateral habenula make it well-suited to impact various

cognitive functions, such as information gathering, flexible reasoning, and spatial memory (Namboodiri et al., 2016). Altering the activity in the lateral habenula interferes with proper defensive responses to stress (Namboodiri et al., 2016). Finally, deep brain stimulation of the lHb reduces cocaine intake and related behaviors (Lax et al., 2013), suggesting the LHb as a potential target for addiction treatment (Fore et al., 2018).

The zebrafish habenula displays spontaneous activity, which is believed to reflect the intrinsic states of the network (Jetti et al., 2014) and may play a role in selecting appropriate behavioral programs with dopamine, acetylcholine, and serotonin release (Bartoszek, 2021). It is shown that this spontaneous activity is not random but has a distinct spatial organization. Moreover, different areas within the habenula exhibit distinct activity patterns, indicating a topographical organization of neural activity (Bartoszek, 2021). Furhermore, Jetti et al. showed that spontaneous activity in the dorsal habenula (dHb) is organized into functionally and spatially distinct clusters of neurons. These clusters are maintained during odor stimulation and play a role in regulating olfactory responses (Jetti et al., 2014). Additionally, studies have shown that the zebrafish dHb appears to play a role in the association phase of classical conditioning, where a neutral stimulus is assigned a negative value (Fore et al., 2018). Furthermore, studies have shown that the zebrafish ventral habenula plays a role in the operant conditioning's goal-directed phase, where the animal learns that a particular behavior offers a chance to escape (Amo et al., 2014).

1.4 Asymmetry in the Habenula

The epithalamus exhibits evolutionary conserved asymmetries in both the pineal complex and the neighboring habenular nuclei (Concha & Wilson, 2001). Differences in how neurons are organized on each side are linked with the asymmetries in size (Concha & Wilson, 2001). All vertebrate classes have some degree of asymmetry in the habenula region (Concha & Wilson, 2001). However, the specific characteristics such as neurochemistry, neuronal organization, and the degree of these asymmetries vary widely (Concha & Wilson, 2001; Ulrike Hüsken, 2014). In frogs and toads, habenular asymmetry is only seen in the dHb and in species like crocodiles and lizards, habenular asymmetry is restricted to medial habenula (Concha & Wilson, 2001). Even though in mammals and birds, the asymmetry of

the habenula is not studied in details, the overall habenula is seen as symmetrical. However, with quantitative volumetric research, a sex-dependent asymmetry in mHb is observed in the chick, and subtle differences in left and right domains are seen in mammals (Concha & Wilson, 2001).

Habenular lateralization is evident through molecular distinctions between the left and right sides, such as uneven distributions of neurotransmitters (Bianco & Wilson, 2009). It is believed that in the nervous system, left-right (L-R) asymmetry is significant for several cognitive processes and behaviors. Habenular asymmetry appears as a L-R difference in the proportion of dorsal habenula (dHb) projection neurons that develop dHbm versus dHbl subnuclear characteristics (Ulrike Hüsken, 2014). The dHbl subnucleus is larger in the left habenula, and the dHbm subnucleus is larger in the right habenula (Ulrike Hüsken, 2014). Additionally, their projections target different regions of the interpeduncular nucleus (IPN) in the ventral midbrain. Specifically, most neurons from the left habenula, which have a dHbl identity, connect to the dorsal part of the IPN (Ulrike Hüsken, 2014). Conversely, most neurons from the right habenula, which have a dHbm identity, connect to the ventral part of the IPN (Ulrike Hüsken, 2014). This results in a lateralized (side-specific) connection pattern where each side of the habenula primarily communicates with different regions of the IPN (Ulrike Hüsken, 2014).

It is shown that the absence of brain lateralization greatly impacts the sensory processing and the functioning of neural circuits. Previous studies on larval zebrafish show that when brain asymmetry is lost, they also lose their responsiveness to visual or olfactory stimuli (Elena Dreosti, 2014). They showed that the sensory responses to visual or olfactory stimuli are lateralized. Specifically, neurons responding to light are located on the left side of the habenula, and neurons responding to odor are more commonly found on the right side (Elena Dreosti, 2014). In addition, they observed that vHb exhibits fewer sensory responses, and these responses were more evenly distributed between the left and right sides (Elena Dreosti, 2014). There are differences in gene expression and neural connections between the left and right sides of the dHb (Elena Dreosti, 2014). These differences include both the output connections (efferent projections) and the input connections (afferent input) (Elena Dreosti, 2014). In addition, the general distributions of glutamatergic, cholinergic, and

peptidergic populations are found to be asymmetric between the two hemispheres (Fore et al., 2018). Moreover, with transgenic lines, it is seen that *brn3a*-expressing neurons are more dominant in the right habenula, while *dsRed2*-positive neurons are predominantly found in the left habenula (deCarvalho et al., 2014). Additionally, several genes, including *kctd12.1*, *kctd12.2, kctd8, adcyap1a*, *nrp1a*, *tac1*, *tac3a*, and *slc5a7a*, exhibit left-right asymmetry in the dorsal habenula (Bianco & Wilson, 2009; Pandey et al., 2018; Powell et al., 2024).

1.5 Single-cell RNA Sequencing

In 2009, single-cell RNA sequencing (scRNA-seq) technology was established and became a popular tool for biological research in different fields (Xie et al., 2021). scRNA-seq provides a detailed, genome-wide profile of RNA molecules within individual cells. This technology allows for comparing individual cells' transcriptomes, which is beneficial for understanding transcriptional similarities and differences among cell groups. Consequently, it facilitates easy analysis of cellular heterogeneity (Haque et al., 2017). scRNA-seq is also helpful for analyzing individual cells, each of which is essentially unique, such as neurons in the brain. Furthermore, scRNA-seq can provide fundamental insights into gene expression, such as identifying co-regulated gene modules based on gene co-expression patterns in single cells (Haque et al., 2017).

1.5.1 Principles of Single-cell RNA Sequencing

The procedures of scRNA-seq primarily involve isolating cells from tissue samples and capturing individual cells, each enclosed with a bead containing unique molecular identifiers within a nanoscale droplet (Dragomirka Jovic, 2022). This step involves extracting high-quality individual cells from tissues to obtain precise genetic data. Cell isolation and capturing methods depend on cell and tissue type. It can be done through entire cell isolation or cell-specific nuclei isolation. As a result of this part, each cell is isolated in a separate reaction mixture, allowing the RNA from each cell to be processed separately. In this reaction mixture, RNA transcripts from the single cell are converted into complementary DNA (cDNA), and each cDNA molecule from a single cell is assigned a unique barcode that acts like a label (Dragomirka Jovic, 2022).

After cDNA is barcoded, it is amplified by either in vitro transcription (IVT) or polymerase chain reaction (PCR). One of the most popular methods for cDNA amplification is PCR amplification with SMART technology (Zhu et al., 2001). In order to limit amplification biases, each individual messenger RNA (mRNA) molecule is barcoded with unique molecular identifiers (UMIs). After single-cell RNA sequencing, the resulting data is analyzed to map and categorize the gene expression landscape within a heterogeneous cell population (Dragomirka Jovic, 2022).

Advantages over Traditional RNA Sequencing Methods and Applications

The single-cell experiments were significantly more conclusive compared to bulk-cell sequencing, which involves sequencing large groups of cells and estimating an average expression level. Unlike bulk sequencing, which prepares libraries from thousands of cells, scRNA-seq generates cell-specific libraries to investigate the DNA and RNA functionalities within distinct cellular subsets (Adil et al., 2021). Hence, traditional transcriptome analyses from bulk RNA/DNA samples only capture the overall signal levels from tissues or organs, failing to identify variations between individual cells (Dragomirka Jovic, 2022). Thanks to its ability to capture single cells, scRNA-seq technology enables the detection of gene expression variability and uncovers cellular heterogeneity within a sample (Trapnell, 2015). Furthermore, since single-cell technologies overcome the limitation of averaging gene expression levels across all cells, increasing the chance of detecting rare cell populations within the tissue (Tang et al., 2009). In neuroscience, scRNA-seq is used to map the diverse cell types in the brain and understand their functions.

1.6 Spatial Transcriptomics

As mentioned in previous subsections, scRNA-seq is a powerful technology that can identify the molecular characteristics of these various cell types by examining the gene expression of single cells extracted from the brain tissue (Dragomirka Jovic, 2022). Despite its effectiveness, scRNA-seq has limitations due to its lack of spatial context. This makes it difficult to fully understand how these cells are arranged and interact to produce specific brain functions (Jung & Kim, 2023).

To overcome these limitations, it is essential to combine scRNA-seq data with spatial information about where each cell type is located within the brain. Spatial transcriptomics map the spatial organization of cell types in the intact tissue. Over the past decade, advancements in imaging and sequencing technologies have led to the development of several spatial transcriptomics tools. These tools have enabled the creation of detailed functional atlases of the brain, which provides deeper insights into the properties of neural circuits (Jung & Kim, 2023).

1.7 Research Objectives and Aim

In this project, I aim to **reveal the molecular topography of vertebrate habenula. To achieve this, I will molecularly define Hb cell types and investigate the topographical organization of these Hb microcircuits within and across Hb hemispheres in zebrafish and mice.**

In this project, I aim to achieve:

- Integrate scRNA-seq datasets with spatial datasets to investigate distinct habenular cell groups, unique gene expressions, and their topographical organization.
- Reveal whether there is a spatial organization of distinct cell types within the habenula and how these cell types are organized.
- Analyze the degree of asymmetry in the habenula of zebrafish and mice by examining the distribution and organization of each cell group.
- Analyze the degree of asymmetry in the habenula of zebrafish and mice by examining the distribution and organization of each cell group, quantify the extent of asymmetry, and provide a quantitative framework to assess hemispheric lateralization as well as its implications for habenular function.
- Analyze the degree of homology between the zebrafish dHb and mice mHb as well as the zebrafish vHb and mice lHb, and investigate the evolutionary conservation between asymmetric and symmetric clusters in zebrafish and mice.

Chapter 2:
MATERIALS & METHODS

All analyses for all datasets were conducted using R as the programming language, with RStudio serving as the integrated development environment (IDE).

2.1 Zebrafish Datasets

2.1.1 Single-cell RNA sequencing dataset

To construct an adult zebrafish scRNA-seq dataset, Pandey et al. dissected entire adult habenula, referencing the expression of the *gng8*-GFP marker, and conducted droplet-based single-cell RNA sequencing (scRNA-seq). The adult zebrafish scRNA-seq dataset accession number is GEO: GSE105115 (Pandey et al., 2018). The section "10X library for adult habenula R1 and R2" is used to download the tpm and count matrices for GSM2818522_adultr1 and GSM2818523_adultr2. With these, the Seurat objects are created. Seurat is a powerful R package developed for quality control, analysis, and exploration of single-cell RNA-seq data (Hao et al., 2024) and used for all of the pre-analysis of datasets used in this study. The first animal data contains 3294 cells, and the second contains 4417 cells.

In order to eliminate the technical or batch effects and increase the robustness and accuracy of the data analysis, these two animals are integrated. Firstly, anchors are identified, which are cell pairs from two datasets that are similar in terms of gene expression profiles (Sharma, 2024). Then, by using these anchors, two animals integrated with Seurat's "IntegrateData" function (Hao et al., 2024), resulting in a dataset with 7711 cells and 40506 genes (Figure X).

Spatial transcriptomics dataset

The Atlas of Zebrafish Transcriptomic Encephalic Cytoarchitecture (AZTEC) dataset is utilized for spatial transcriptomics analysis of adult zebrafish. Although I was not involved in the construction of this dataset, I have been granted access to the processed data from the habenulae region only. The full dataset remains unpublished, with its main article

currently in the manuscript phase (Bredesen-Aa, 2024). Therefore, in this methodology section, I will outline the general methodology for constructing AZTEC.

Adult zebrafish spatial transcriptomics data is constructed with Resolve Biosciences' "The Molecular Cartography Platform" (Bredesen-Aa, 2024). It is an innovative multi-analyte, highly multiplex spatial analysis technology. This platform generates comprehensive contextual datasets that reveal molecular interactions at a subcellular level and maintain the integrity of the sample tissue. Resolve Biosciences' Molecular Cartography Platform offers a high-resolution view of transcriptomic activity and enables the analysis of hundreds of genes in a single experiment (Benedict Nilges Resolve Biosciences). Six-month-old zebrafish were euthanized in cold water, and their brains were dissected and fixed following Resolve guidelines. Brains were incubated in PAXgene® fixation solution, stabilized in PAXgene® tissue stabilizer, and cryoprotected in a sucrose solution before being cryo-embedded and sectioned (Percival P. D'Gama, 2021). Sections were collected on precooled coverslips and sent to Resolve GmbH, Germany, for Molecular Cartography analysis using Paxgene-fixed samples. The analysis followed the manufacturer's instructions, involving probe hybridization, imaging, and color development to decode gene-specific signals (Percival P. D'Gama, 2021). Molecular Cartography sections were aligned to the Adult Zebrafish Brain Atlas (AZBA) (Kenney et al., 2021) using custom software (Bredesen-Aa, 2024). Then, using Cellpose (Stringer et al., 2021) on DAPI stains, nuclei have been depicted. Nuclear transcriptomes have been determined by assessing the transcript enrichment for each gene within individual nuclei. The spatial positions of nuclei have been determined by calculating the centroids of the depicted nuclei (Bredesen-Aa, 2024).

I received the Seurat object of the habenula-only AZTEC dataset, which includes samples from two 6 month adult zebrafish, labeled animal 1 and 2. In the original complete AZTEC dataset, the samples from different animals were not integrated because no batch effect was observed, allowing them to be merged directly (Bredesen-Aa, 2024). To maintain consistency with the original data and due to the lack of batch effects, I did not perform any further integration for this AZTEC habenula dataset. AZTEC data for habenula includes 11,384 cells and 99 genes in total (Figure 3.10).

2.2 Mice Datasets

2.2.1 Single-cell RNA sequencing dataset

To download the adult mice scRNA-seq dataset, the link https://allen-brain-cell-atlas.s3.us-west-2.amazonaws.com/index.html is used. Here, the dataset used in this project is accessible through: **allen-brain-cell-atlas -> expression_matrices -> WMB-10Xv3 -> 20230630 -> WMB-10Xv3-TH-raw.h5ad**. In this context, "TH" refers to the thalamus. The file was downloaded because the researcher pointed out that the medial habenula (mHb) and lateral habenula (lHb) are included in the TH–EPI-Glut groups (Yao et al., 2023).

Due to the scope of this project being limited to the habenula, the thalamus dataset was subsetted to include only the cells labeled '08 MH-LH Glut' in the 'class' section of the metadata and seurat object is created with this subsetted data. This includes 6,053 cells and 32,285 genes, derived from two animals; one female and one male. Since these two animals were already integrated, no further integration was performed (Figure 3.6).

Spatial transcriptomics dataset

The spatial transcriptomic dataset for adult mouse was constructed using multiplexed error-robust fluorescence in situ hybridization (MERFISH) (Zhang et al., 2023). This dataset is available for visualization in the Allen Brain Cell Atlas.

To download the adult mouse MERFISH dataset, the following link is used: Allen Brain Cell Atlas. The specific dataset for this project can be accessed via: **allen-brain-cell-atlas -> expression_matrices -> MERFISH-C57BL6J-638850-> 20230830 -> C57BL6J-638850-raw.h5ad.** The MERFISH dataset is subsetted by the “17 MH-LH Glut” labeled cells in the 'class' section of the metadata, and a Seurat object is created accordingly. This subset includes 5759 cells and 550 genes, specifically annotated as glutamatergic habenular neurons. Two sections of this dataset included only one cell each, while another section contained two cells. These sections, along with two additional sections, were removed from the dataset as their cells were dispersed outside the habenula. Additionally, 50 out of the 550 genes were labeled as “blank” and were not present in the MERFISH whole mouse brain

atlas (Zhang et al., 2023). The resulting filtered MERFISH habenula dataset consists of 5714 cells and 500 genes. Since this dataset only includes one animal, no integration is needed.

2.3 Normalization & Variance Stabilization

Before further analysis, it is crucial to normalize and apply variance stabilization since the technical variability in the data occurred from differences in sequencing depth and other factors, which can result in misleading interpretations and conclusions about gene expression levels of the cells across the dataset (Choudhary & Satija, 2022).

In more detail, due to differences in lysis and reverse transcription efficiency, fluctuations in cellular RNA content, and random sampling during next-generation sequencing, variations in cellular sequencing depth are likely to arise in scRNA-seq datasets. Data normalization is required to adjust these differences. In addition, the genes expressed in the datasets I used vary in expression levels and abundance due to the nature of the cells. This variance in sequencing depth remains a significant source of technical variance (Choudhary & Satija, 2022). Hence, variance stabilization is needed to ensure that both lowly and highly expressed genes can equally influence the downstream definition of cellular state. Furthermore, scRNA-seq data is likely to have overdispersion (the variance exceeds the mean) (Choudhary & Satija, 2022).

Since sctransform uses regularized negative binomial regression, it is more suitable for datasets with overdispersion. For these reasons, I used stransform v2, which is the default version in the Seurat workflow for all the datasets in this project. This method uses Pearson residuals from negative binomial regression as input for standard dimensional reduction techniques. It is shown that sctransform v2 normalization allows for efficient differential analysis across datasets. Moreover, it reduces false-positive results, even with significant differences in sequencing depth across experiments, and enhances performance in crucial downstream analyses, such as variable gene identification and differential expression (Choudhary & Satija, 2022). Details for this updated sctransform v2 can be found in the manuscript published by Choudhary and Satija.

The next important step is deciding the variable genes across the dataset because, as the name implies, these genes indicate strong biological variation and are key factors for identifying cellular heterogeneity and cell identity (Choudhary & Satija, 2022). By default, sctransform chooses 3000 variable genes; however, this gene number is too many for our datasets with lower variable gene numbers. In order to decide the variable gene number used in sctransform, a visualization method called variable feature plot, implemented in the Seurat package, is used. This method utilizes the modeling of the mean-variance relationship. The variance of standardized values for each gene is calculated, which indicates the dispersion measure after accounting for mean expression (Stuart et al., 2019). This measure is used to rank the genes, and details can be found in the article by Stuart et al.

By utilizing this method, 60 out of 99 genes were selected as variable genes in the AZTEC dataset (Figure 3.10). In the MERFISH dataset, 300 out of 500 genes were chosen as variable genes (Figure 3.19). For the adult zebrafish scRNA-seq dataset, 1000 variable genes were selected (Figure 3.2), while 900 variable genes were chosen for the mouse scRNA-seq dataset (Figure 3.6).

2.4 Principal Component Analysis

Principal component analysis (PCA) is a statistical method that selects the significant patterns and trends of variance in the multidimensional data to transform a large set of most informative variables into a smaller set of principal components (PCs). These principal components are where the variance is highest (Greenacre et al., 2022). For instance, the first principal component captures the most variance; the second principal component captures the second most variance independently of the 1st.

PCA begins by standardizing the data and converting all features to a common scale. Next, it computes the covariance matrix to reveal the relationships between variables. From this matrix, eigenvectors and eigenvalues are derived, where eigenvectors identify the principal components and eigenvalues indicate their importance. Finally, the data is transformed into the new space defined by these principal components (Greenacre et al., 2022; Jolliffe, 2002).

This approach allows PCA to effectively reduce dimensions while retaining as much essential information as possible. In addition, by focusing on high variance, PCA can filter out the fluctuations or noise that are insignificant for the analysis, making it cleaner and more interpretable (Greenacre et al., 2022; Jolliffe, 2002). Furthermore, PCA allows us to reduce the data to 2 or 3 principal components, making it possible to plot and visually examine the data (Van der Maaten, 2008). Overall, PCA simplifies dataset exploration, visualization, and downstream analysis (Greenacre et al., 2022).

In order to choose optimal PC numbers for the datasets used in this project, I utilized the ElbowPlot function from the Seurat package. The elbow plot ranks the PCs by their percentage of variance and generates a plot with the PCs' standard deviations. In this plot, the elbow point aligns well with the significant dimensions, which indicates the optimal PC (Butler, 2018). Based on elbow plots for each dataset, 30 PCs were selected as the threshold for the scRNA-seq mice dataset (Figure 3.6) and scRNA-seq zebrafish datasets (Figure 3.2) and 26 PCs were selected as the threshold for MERFISH (Figure 3.19). For AZTEC spatial datasets, PCA was not used due to the limited number of genes (99 genes in total), which hampers the efficiency of PCA (Lenz et al., 2016).

2.5 Finding Neighbours and Clusters

To create distinct cell groups and identify the similarities as well as the distances, the Seurat package workflow is used (Hao et al., 2024). With the FindNeighbors function implemented in Seurat, a k-nearest neighbors (KNN) graph is constructed. This graph refines the edge weights between any two cells based on the shared overlap in their local neighborhoods and based on Euclidean distance in previously determined PCA space (Butler, 2018). I utilized this function with pre-determined PCs for scRNA-seq datasets from mouse and zebrafish, as well as the spatial dataset MERFISH. However, a different approach is required for AZTEC since PCA is not used during preprocessing. For this dataset, instead of selecting PC values, all genes are directly used as features in the FindNeighbors function, meaning the genes in the dataset are used to calculate the nearest neighbors (Bredesen-Aa, 2024).

Then, with the FindClusters function, which is also implemented in the Seurat package, the Louvain community detection method is applied for modularity optimization (Stuart et al., 2019). It iteratively groups cells together and optimizes the standard modularity function (Stuart et al., 2019). This function includes a parameter called resolution, which determines the granularity of the clustering.

2.6 Choosing optimal resoluiton with Clustree

The resolution value is directly proportional to the number of identified clusters; higher resolution values generally result in a larger number of smaller, more specific clusters. However, a higher resolution does not necessarily mean more meaningful or biologically relevant results. The optimal resolution should be determined carefully, balancing detailed cluster granularity with biological interpretability and relevance.

Considering the importance of selecting the optimal resolution value, a visualization tool called "clustree" which is short for "clustering trees" is used. It provides a tree diagram showing resulting clusters at different resolution values. With this tool, information about cluster stability and cells that change their clusters in different resolutions can be identified (Luke Zappia & Oshlack, 2018).

I constructed Clustree diagrams for all datasets using different resolution values ranging from 0.1 to 5. Considering cluster stability and the division of clusters, I determined the optimal resolution values as follows: 1.5 for scRNA-seq zebrafish data, 1.5 for AZTEC zebrafish data, 1.8 for scRNA-seq mice data and 1.7 for MERFISH mouse data. The Clustree plots for each dataset are provided in the results section (Figure 3.3, Figure 3.7, Figure 3.11, Figure 3.20).

2.7 Taxonomy and Visualization

The ordering of clusters based on their molecular similarities is done with a function in CoCy that calculates a cluster taxonomy by hierarchical clustering of centroids (Bredesen-Aa, 2024). In detail, for each cluster, this function computes the centroid, which is the average gene expression profile of all the cells within that cluster. Once the centroids for all clusters are calculated, these centroids are compared to each other. Then, hierarchical

clustering is applied to these centroids. Lastly, a dendrogram that displays the arrangement of clusters formed during hierarchical clustering is created with CoCy (Bredesen-Aa, 2024).

Uniform Manifold Approximation and Projection (UMAP) is a dimensionality reduction technique grounded in the principles of Riemannian geometry and algebraic topology (McInnes et al., 2018). It is commonly used for visualizing high-dimensional data by projecting it into lower dimensions (McInnes et al., 2018). In this study, UMAP is used to visualize the identified clusters in two dimensions (Figure 3.4). The FindMarkers function from the Seurat package is used to visualize gene expressions for each cluster. The expression profiles for the marker genes of each cluster are then visualized using the DoHeatMap function within Seurat. This function generates a heatmap of single-cell gene expression data, utilizing ggplot2 for visualization (Figure 3.4) (Butler, 2018).

In order to see the relationship between the spatial proximity of points and the similarity of their gene expression profiles, a correlation versus distance plot is generated with CoCy (Bredesen-Aa, 2024). The Pearson correlation coefficient of their gene expression profiles is calculated with the function for each pair of spatial points. This measure quantifies the similarity in gene expression between two points, ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation). Then, with CoCy, the pairwise correlation coefficients are plotted against their corresponding distances in the dataset (Bredesen-Aa, 2024).

Spatial images for AZTEC and MERFISH (Figure 3.13, Figure 3.22) are created with a plot function in CoCy (Bredesen-Aa, 2024). This function uses the common coordinate framework (CCF) coordinates (Wang et al., 2020) of MERFISH and cell position coordinates of AZTEC to draw anatomical locations of clusters in 2D from multiple views which are horizontal, coronal and sagittal (Bredesen-Aa, 2024). In a similar manner, gene expression plots for AZTEC (Figure 3.17) are created with a CoCy function that draws anatomical distributions of a feature/gene in 2D in the coronal section (Bredesen-Aa, 2024).

2.8 Dataset Integration for Comparison

Integrating these datasets is essential to comparing mouse and zebrafish data and enhancing the dataset size and gene pools of the spatial datasets MERFISH and AZTEC. In order to do this, an unpublished R package named CoCy is used and developed in the Yakşi lab (Bredesen-Aa, 2024). First, the EggNOG map is created for mouse and zebrafish. EggNOG stands for evolutionary genealogy of genes: non-supervised orthologous group. It is a database that provides orthologous groups of genes from various species, offering robust functional annotations and enabling high-throughput analyses of orthologous relationships between genes from different genomes (Muller et al., 2010). EggNOG-mapped versions of the Seurat objects are created with the EggNOG mapping code that is included in CoCy (Bredesen-Aa, 2024)

CoCy package offers different integration methods; for this project, datasets integrated with the CoCy package utilizing Seurat anchor-based Canonical Correlation Analysis (CCA) integration and the K Mutual Ratio (KMR) calculation of similarity among clusters of two datasets across their joint embedding (Bredesen-Aa, 2024). With KMR, the overlap between two datasets is quantified. Firstly, the k nearest neighbors (KNNs) between the datasets are identified. For a sample in dataset A, its KNNs in dataset B are found. Then, the KNNs are filtered to the samples in B that also have at least one KNN in A with the same cluster label. These samples are counted as mutual neighbors. For a cluster label lA of A and lB of B, mA denotes the number of samples in A with label lA that has at least one mutual neighbor with label lB. Similarly, mB is calculated by counting the samples in dataset B whose KNN in A has a KNN in B with the same label. The KMR is the sum of mA and mB, providing a measure of the mutual overlap between the clusters in the two datasets. Calculating the KMR for each combination of cluster labels and normalizing the resulting matrix produces a matrix of ratios of matches between clusters. This matrix helps to understand how well the samples in each dataset correspond to each other based on their nearest neighbors (Figure 3.16), (Bredesen-Aa, 2024).

2.9 Cluster Asymmetry & Lateralization Index Calculation

Spatial AZTEC and MERFISH datasets were used to calculate the asymmetry percentages of the cells in each cluster of the habenula. First, cells on the left and right sides

of the habenula were separated based on coordinates. The number of cells on the right and left sides was calculated for each cluster. Then, the percentage of cells on each side relative to the total number of cells in the cluster was calculated.

The lateralization index formula measures the degree of asymmetry between two groups of cells positioned on the right and left sides of the habenula. The formula is given by the absolute value of the difference between the number of right cells and left cells, divided by the total number of cells (the sum of right and left cells) (Equation 2.1). This calculation normalizes the asymmetry, producing a value between zero and one. A value of zero indicates perfect symmetry (equal numbers of right and left cells), while a value of one indicates complete asymmetry (all cells on one side and none on the other).

$$LI = \left|\frac{(\#\ of\ right\ cells)-(\#\ of\ left\ cells)}{(\#\ of\ right\ cells)+(\#\ of\ left\ cells)}\right| \quad (2.1)$$

Lastly, lateralization of genes in AZTEC and MERFISH are calculated to have a comprehensive understanding of gene expression asymmetry in the habenula of zebrafish and mouse. First, the left and right cells are separated by the coordinates of the cells. Then, I extracted the counts for the left and right cells, removing any counts present in fewer than 30 cells. In the AZTEC dataset, 73 genes are found to be present in more than 30 cells. In the MERFISH dataset, 305 genes are found to be present in more than 30 cells. Next, the sum of counts for each gene in both left and right cells is calculated. Then, the lateralization of genes is calculated based on the formula (Equation 2.2). Instead of using the number of cells, I used the counts for each gene in both left and right cells. Since the MERFISH dataset includes 305 genes, the top 60 lateralized genes are selected for visualization (Figure 3.31). In the AZTEC dataset, all 73 genes are shown in the plot (Figure 3.31). The visualization graph is constructed with the ggplot R package.

$$LI = \left|\frac{(\#\ of\ right\ counts)-(\#\ of\ left\ counts)}{(\#\ of\ right\ counts)+(\#\ of\ left\ counts)}\right| \quad (2.2)$$

2.10 HCR & Clearing of Samples

The hybridization chain reaction (HCR) is utilized for in situ hybridization, providing isothermal, enzyme-free signal amplification. This technique enables multiplexed, quantitative, and sensitive imaging of mRNA expression. HCR is particularly advantageous in challenging imaging scenarios, such as whole-mount vertebrate brains and thick tissue sections, by amplifying the signal above background autofluorescence (Choi et al., 2018). In this project, HCR is used for whole-mount adult zebrafish brain and juvenile zebrafish brain imagining with two genes of interest, *aoc1*, and *grm6a*. I combined HCR with the clearing protocol from Binaree since tissue optical clearing allows for extensive volume imaging and data analysis without damaging the tissue (Lee et al., 2022). The following protocol for HCR is created based on the chapter written by D'Gama et al. (D'Gama & Jurisch-Yaksi, 2023) and combined with the modified clearing protocol by Binaree (Binaree).

Two adult animals were euthanized in cold water by a lab technician certified in animal experimentation. Immediately after euthanasia, the brains were dissected based on the protocol (Jeong et al., 2022) and transferred into a 1.5 mL tube containing 200 µL of 4% PFA in a fume hood for overnight incubation at 4 °C. The next day, the samples were washed with 1X dPBS while shaking at room temperature for 5 minutes, repeated thrice. Then, the samples were dehydrated and permeabilized with a series of methanol (MeOH) washes (1 mL each): (a) 100% MeOH for 4 x 10 minutes and (b) 100% MeOH for 1 x 50 minutes. After this step, the samples were stored at -20 °C overnight. The next day, the samples were rehydrated with a series of graded 1 mL MeOH/dPBS washes for 5 minutes each at room temperature: (a) 75% MeOH in dPBS, (b) 50% MeOH in dPBS, (c) 25% MeOH in dPBS, and (d) dPBS.

To clear the tissue, the initial solution from Binaree was used to incubate the samples at 4 °C overnight until they sank. The samples were then transferred into 500 µL of Tissue Clearing Solution B, wrapped in aluminum foil to protect them from light, and incubated in a water bath at 37 °C for 24 hours. The next day, the samples were washed with reverse osmosis (RO) water while shaking at 30 rpm at 4 °C for 30 minutes, repeated four times.

For the detection stage, samples are pre-hybridized with 500 µL probe hybridization buffer for 30 minutes at 37°C. After removing the probe hybridization buffer, a probe solution (2 pmol of each probe set, prepared by mixing 2 µL of 1 µM stock in 500 µL of

probe hybridization buffer at 37°C) is added. The samples are then incubated overnight (12–16 hours) at 37°C. The next day, the samples are washed four times for 15 minutes each with 500 µL of probe wash buffer at 37°C to remove excess probes. Finally, the samples are washed twice for 5 minutes each with 500 µL of 5x SSCT at room temperature.

For the amplification stage, samples are pre-amplified with 500 µL of amplification buffer for 30 minutes at room temperature. During this time, 30 pmol of hairpin h1 and 30 pmol of hairpin h2 are separately prepared by snap cooling 10 µL of 3 µM stock. The tubes are heated to 95°C for 90 seconds and then cooled to room temperature in a dark drawer for 30 minutes. HCR hairpins h1 and h2 are provided in the hairpin storage buffer, ready for snap cooling, and should be snap-cooled in separate tubes.

After this, the hairpin solution is prepared by adding the snap-cooled h1 and h2 hairpins to 500 µL of amplification buffer at room temperature. The pre-amplification solution is then removed, the hairpin solution is added, and the samples are incubated overnight (12–16 hours) in the dark at room temperature.

Lastly, excess hairpins are removed by washing with 500 µL of 5x SSCT at room temperature as follows: (a) 2 x 5 minutes, (b) 2 x 30 minutes, (c) 1 x 5 minutes, and with dPBS for 3 x 5 minutes. The samples are then transferred to 500 µL of Mounting Solution from Binaree and incubated overnight (12–16 hours) in the dark at room temperature.

2.11 Confocal Imagining

In this stage, the samples become transparent and difficult to see with the naked eye. I carefully remove the samples from the mounting solution and place them onto a slide using a Pasteur pipette (with the tip slightly cut to prevent damage to the adult brain). The brain is then positioned with a microinjection pipette tip to ensure a symmetric orientation, with the dorsal part facing upward.

Since this is whole-brain imaging, I avoid squishing the brain by using a glass slide with a cover slip (18 x 18 mm) bridge made using glue (D'Gama & Jurisch-Yaksi, 2023). I used five coverslips for the samples to create the cover slip bridge. Finally, a longer coverslip (24 x 60 mm) is placed on the coverslip bridge. LSM880 upright Zeiss confocal microscope

with 20X (Numerical Aperture: 0.8, Working Distance: 0.55 mm) objectives is used for imaging the samples. The HCR amplifier fluorophore channels were 647 for *grm6a* and 546 for *aoc1*. For *aoc1*, the master gain is set to 680, the laser power to 60%, and the excitation filter wavelength ranges from 553 to 597 nm. For *grm6a*, the master gain is set to 700, the laser power to 80%, and the excitation filter wavelength ranges from 652 to 758 nm. For both genes, an averaging factor of 4 was applied to improve image quality.

Chapter 3:

RESULTS

3.1 Analysis of Zebrafish Habenula Single-cell RNA Sequencing Dataset Revealed Multiple Distinct Cell Types of Adult Zebrafish Habenula.

Anatomical regions of the zebrafish habenula based on architecture, neuronal connectivity, and neurotransmitter identity have been studied in previous research (Schmidt & Pasterkamp, 2017). As mentioned in the introduction, the zebrafish habenula subdivides into dorsal (dHb) and ventral (vHb) domains (Figure 3.1) (Schmidt & Pasterkamp, 2017). Previous studies report different gene markers and cell groups within the zebrafish habenula utilizing histology (Amo, 2010; deCarvalho et al., 2014; Hong et al., 2013; Lekk et al., 2019; Turner et al., 2016) and single-cell RNA sequencing methods (Pandey et al., 2018).

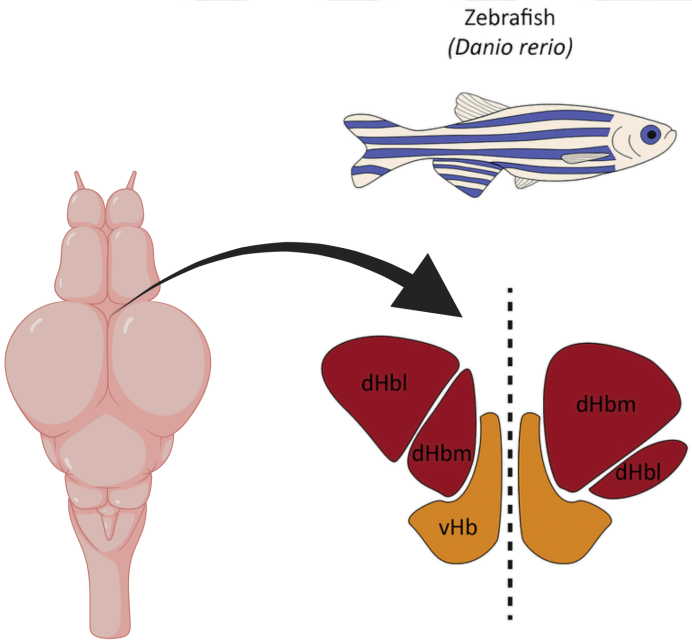

Figure 3.1: Anatomical organization of the zebrafish habenula. The dorsal view of the zebrafish brain (left), an anatomical illustration of adult zebrafish habenula, divided into its subregions: the dorsal habenula lateral (dHbl), dorsal habenula medial (dHbm), and ventral habenula (vHb) (Schmidt & Pasterkamp, 2017), (Created with BioRender).

To analyze the various groups of neuronal cells in the zebrafish habenula, I utilize the adult zebrafish scRNA-seq dataset provided by Pandey et al. (Pandey et al., 2018). This dataset comprises habenular neurons from two adult zebrafish.

Two datasets of these animals might differ in terms of technical or batch effects. Also, integrating datasets provides increased sample size and increases the robustness and accuracy of the data analysis. That is why these two animals are integrated (see methods), resulting in a dataset that includes 7711 cells and 40506 genes (Figure 3.2c). Before further analysis, I normalized and stabilized the variance caused by technical differences like sequencing depth to prevent misleading conclusions about gene expression. I used sctransform, which applies regularized negative binomial regression and is effective for datasets with overdispersion, where variance is higher than the mean (Choudhary & Satija, 2022). The next important step is selecting variable genes across the dataset because, as the name implies, these genes indicate strong biological variation and are key factors for identifying cellular heterogeneity and cell identity (Choudhary & Satija, 2022). To select highly variable genes, I utilized a method implemented in the Seurat package (see methods) and chose 1000 as the highly variable genes value for sctransform (Figure 3.2a). Next, I performed principal component analysis (PCA), and I chose the PC number as 30 based on the standard deviation and elbow plot for each PCs (see methods) (Figure 3.2b).

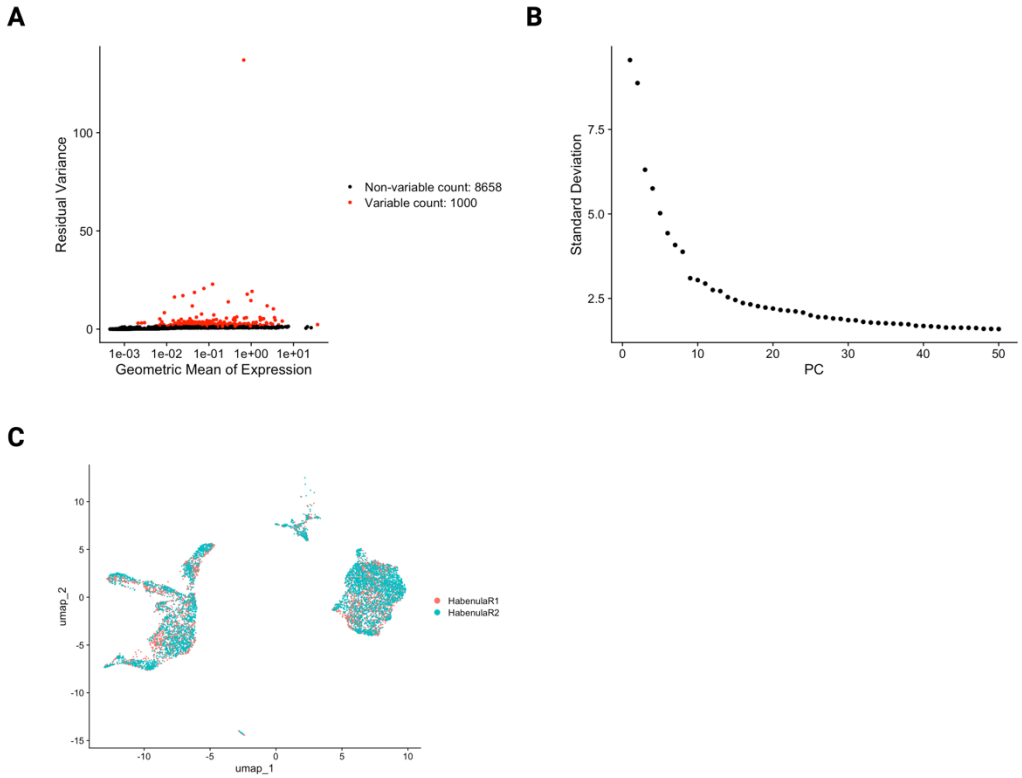

Figure 3.2: Analysis of the variable genes, PCA, and cell distribution of animals in the scRNA-seq zebrafish data. (A) Scatter plot (Variable Feature Plot) depicting the residual variance as a function of each gene's geometric mean of expression. Genes are categorized into variable (red dots, 1000 genes) and non-variable (black dots, 8658 genes) based on their variance. Variable genes show higher residual variance compared to non-variable genes. (B) Elbow plot showing the standard deviation of the principal components (PCs) derived from the PCA of the gene expression data. The first few PCs capture most of the variance, with a steep drop-off indicating the primary components of variation in the dataset. From this plot, the 30th principal component was selected for further analysis of the dataset. (C) UMAP plot illustrating the distribution of cells from two different animals integrated into the zebrafish scRNA-seq dataset (animal 1 in red and animal 2 in cyan). The plot demonstrates the successful integration of cells from both animals.

In order to determine the number of transcriptionally distinct clusters in this dataset, it is essential to select the appropriate resolution value, a critical parameter in Seurat's clustering functions. The graph-based clustering algorithm in the Seurat package created a cell similarity-based shared nearest neighbor (SNN) graph, and the resolution value affects how clusters are cut in the SNN graph. Considering the importance of selecting optimal resolution value, I utilized a visualization tool called clustree, which is short for “clustering trees” (Luke Zappia & Oshlack, 2018). It provides a tree diagram that shows the resulting clusters at different resolution values. This tool allows for the identification of cluster

stability and the cells that jump between the clusters across different resolutions (Luke Zappia & Oshlack, 2018). I constructed a clustree diagram of the adult zebrafish habenula data with different resolution values between 0.1-3, 4, and 5 (Figure 3.3a). Then, for all different resolution values, I created Uniform Manifold Approximation and Projection (UMAP), which simplifies large datasets by reducing their dimensions but keeps both the global and local structural nuances, making it easier to understand and visualize the data (McInnes et al., 2018). Figure 3.3b shows three examples of these UMAPs with different resolution parameters, which are 0.5, 1.5, and 3. Considering cluster stability, the division of clusters, their UMAPs, and the distribution of gene expressions for each cluster, I decided that a resolution of 1.5 is optimal for my study as it provides distinct and stable cell groups. My analysis at 1.5 resolution resulted in 22 transcriptionally distinct cell groups (Figure 3.3b, middle).

Cluster 14 lacks expression of *snap25* (synaptosome-associated protein 25a), a key neuronal marker, but does express *hepacam* (hepatic and glial cell adhesion molecule) (NCBI, 2024; Sjostedt et al., 2020). Moreover, this cluster exhibits a robust expression of the "Her family" genes *her15.2*, *her4.2*, *her4.3*, and *her4.4*, typically observed in neuronal progenitor cells. These genes are instrumental in maintaining the undifferentiated state of progenitor cells or in regulating the timing of their maturation into neurons (Liu, 2022). The scope of this project focuses on habenular neuronal cells. Moreover, non-neuronal cells are scarce in the fish dataset used in this project, and the mice dataset does not include any non-neuronal cells. Therefore, I decided not to include non-neuronal cells in this study. This removal resulted in 21 transcriptionally distinct clusters that can be visualized in UMAP (Figure 3.4a).

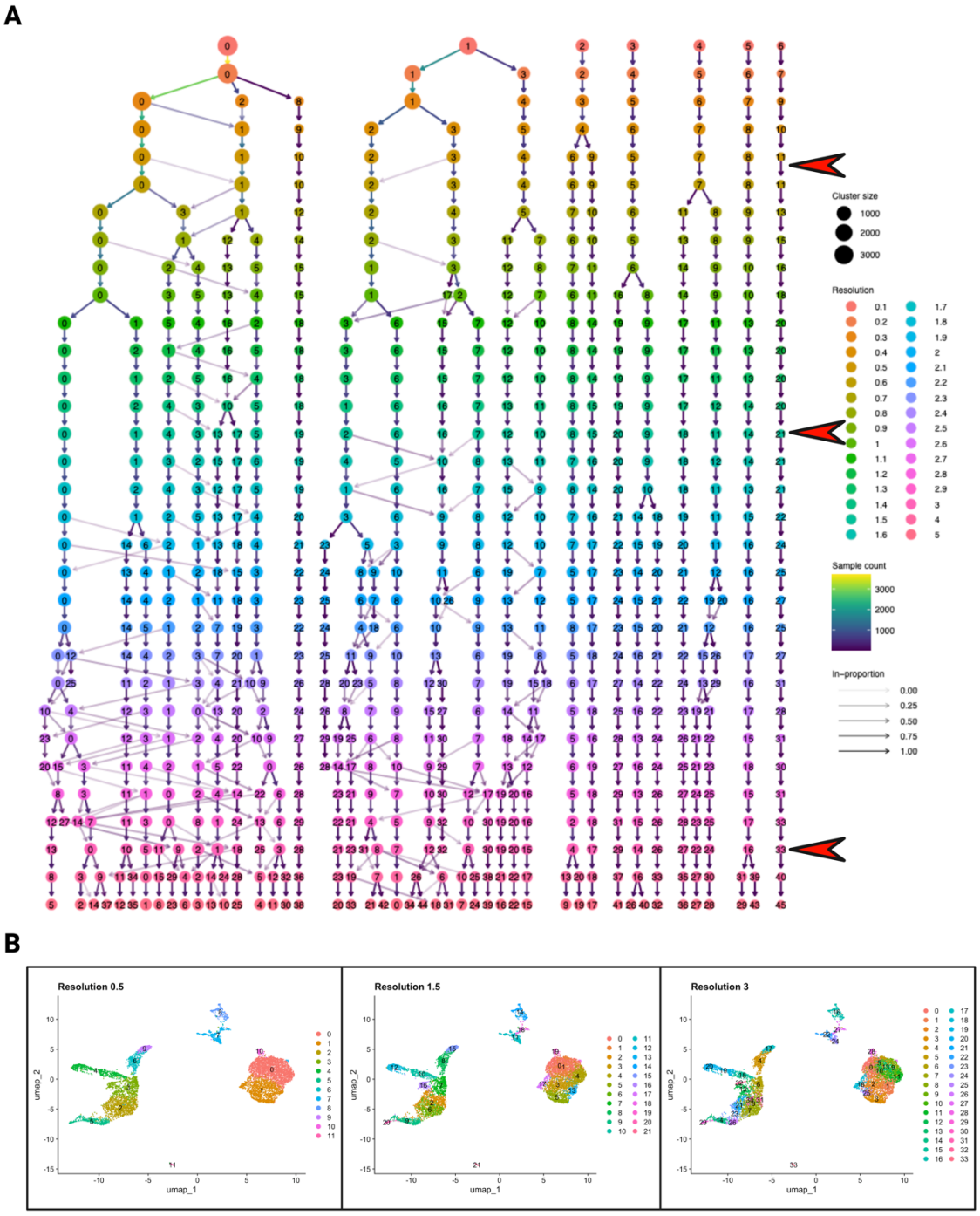

Figure 3.3: Cluster tree and UMAPs of the scRNA-seq zebrafish data. (Top) Cluster tree for resolutions between 0.1 and 5; nodes are colored based on the value of resolution. Cluster size is indicated by the size of the nodes. Edges are colored to reflect the sample size, ranging from blue (fewer samples) to yellow (more samples). The transparency of the edges indicates the significance of the connection. Red arrows on the side indicate selected resolutions 0.5, 1.5, and 3 as examples. (Bottom) Each UMAP illustrates how the number and composition of clusters vary with different resolution settings (Created with BioRender).

Next, I created a dendrogram of these clusters based on their molecular similarity in order to differentiate and understand the differences and similarities among cell groups. Then, I created a heatmap with the top expressed genes of each cluster (Figure 3.4b). Cluster 1 exhibits high and specific expression of the *tubb5*, *tmsb*, and *vim* genes. Cluster 2 shows strong and specific expression of *gad2*, *gad1b*, and *slc32a1*. *Pvalb7*, *capgb*, and *cxcl14* are highly specific to Cluster 3, a small cell group comprising 40 cells. The *spa17* gene shows significantly high expression in Cluster 3 but is not specific to it, as it also exhibits high expression in Cluster 12. Cluster 3 is also distinct from the rest of the data, as illustrated in the UMAP (Figure 3.4a). Cluster 4 expresses the *cbln2b* gene. Cluster 5 exhibits a significantly high expression of *gng2*, although *gng2* is also seen in clusters 6, 8, and 10, with the highest level in cluster 5. Cluster 6 shows expression of *lypd6*, *msi1*, *drgx*, *gpr78a*, *wnt7aa*, *camta1a*, and *mad1l1*. These genes also appear in other dorsal-like clusters but are most highly expressed in cluster 6. *Si:dkey-159a18.1* is specific to cluster 7, which is one of the three clusters with high *sst1.1* expression. Cluster 7 shares expression of *rally*, *lrrtm1*, *im:7152348*, and *pouf3f1* with cluster 8. Additionally, cluster 8 expresses *gng2*, which is not seen in cluster 7. The dendrogram indicates that clusters 7 and 8 are molecularly similar, with gene expression differences in *sst1.1*, *gng2*, and *Si:dkey-159a18.1*. *Tac3a* and *g0s2* genes exhibit particularly specific expression in clusters 9, 10, and 11, with the highest levels in cluster 9. The *impact*, *fxyd1*, and *murcb* genes are expressed only in clusters 9 and 11. Additionally, the *nrgna* gene is highly specific to clusters 9, 10, 11, and 13. *Sst1.1* is highly expressed in clusters 7, 11, and 12. Besides *sst1.1* and *spa17*, cluster 12 shows high expression of *pyyb*; however, these genes are not specific to cluster 12 and are expressed in other clusters as well. The only genes specific to cluster 12 are *trh* and *spx*.

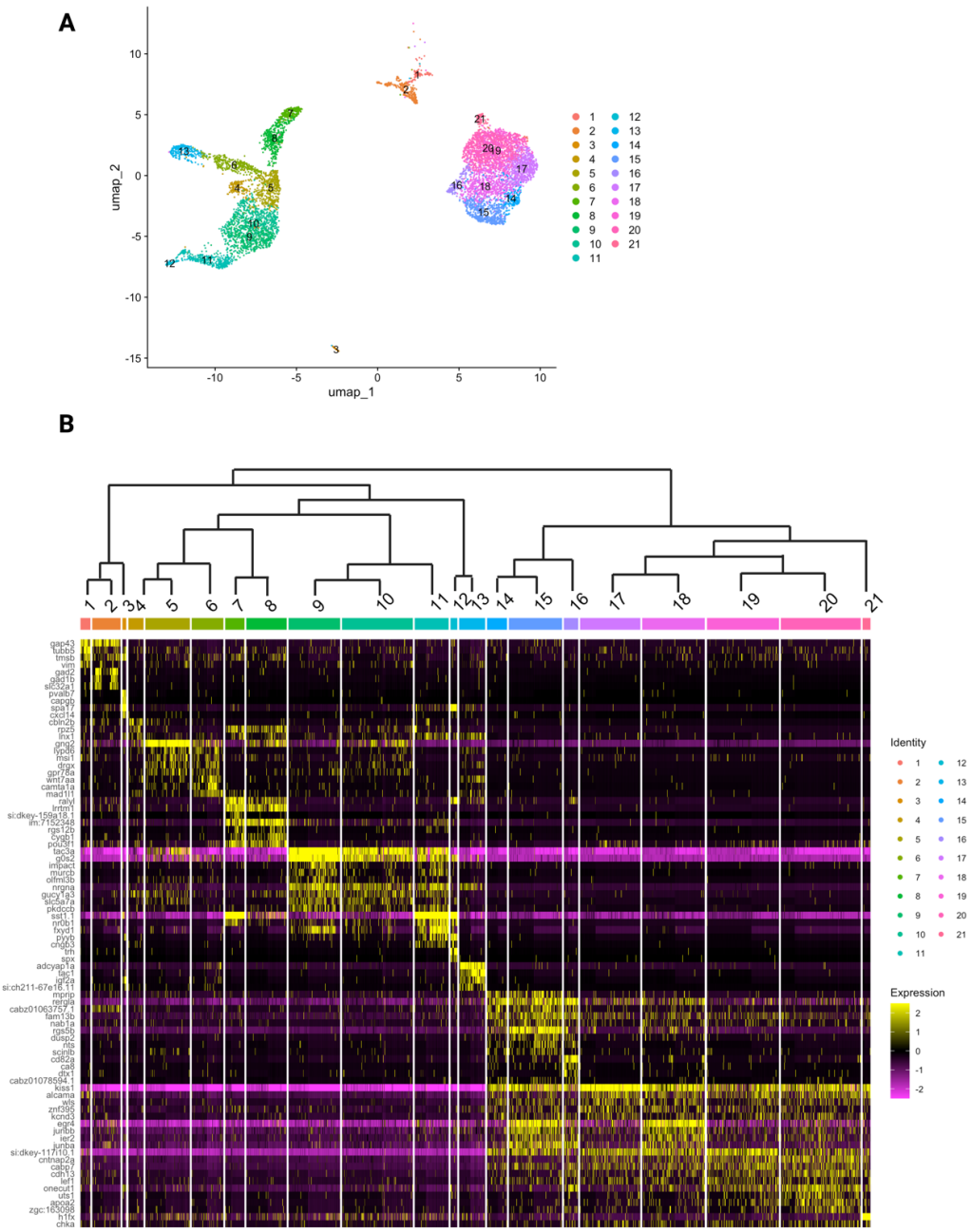

Figure 3.4: UMAP and the heatmap of zebrafish scRNA-seq data with dendrogram. (A) UMAP visualization of adult zebrafish scRNA-seq data, showing clusters sorted by molecular similarity. (B) Heatmap depicting gene expression profiles with an accompanying dendrogram illustrating the hierarchical clustering of the cells. Rows represent gene names, and columns represent clusters. The color intensity in the heat map indicates gene expression levels, with yellow indicating higher expression and black indicating no expression.

Adcyap1a, *tac1*, *igf2a*, and *si:ch211-67e16.11* are expressed exclusively in cluster 13. Notably, clusters 14 through 21 show high expression of *kiss1* and *aoc1*, genes predominantly expressed in the ventral region (Pandey et al., 2018), suggesting these clusters are in the ventral habenula. Additionally, Pandey et al. identified *mprip*, *cd82a*, and *cntnap2a* as ventral differential markers (Pandey et al., 2018). Specifically, *mprip* is highly expressed in clusters 14, 15, and 16; *cd82a* is exclusive to cluster 16; and *cntnap2a* is present in clusters 17 through 21. Clusters 15 and 18 also show strong and distinct expression of immediate early genes *egr4*, *ier2*, *junbb*, and *junba*, which are linked to synaptic plasticity (Mech et al., 2021; Norton, 2019). Furthermore, *H1fx* is uniquely specific to cluster 21. As illustrated in the heatmap, aside from these genes, other ventrally expressed genes such as *alcama*, *wls*, *znf395*, *kcdn3*, *cabp7*, and *lef1* do not exhibit cluster specificity but instead show a general expression profile among ventral clusters (Figure 3.4).

Notably, the expressed genes in ventral clusters (14-21) exhibit a more scattered profile, with all genes being expressed in similar patterns within these clusters, with a few exceptions. Conversely, the dorsal clusters (3-13) demonstrate a more distinct separation of gene expression, indicating that the dorsal cell types might possess greater molecular heterogeneity compared to ventral cell types. Clusters 1 and 2, which are distinct from other dorsal and ventral cell groups, as shown in Figure 3.4a, do not express typical dorsal or ventral gene markers. These clusters' anatomical locations will be predicted when this dataset is compared with the spatial dataset, AZTEC.

3.2 Comprehensive Analysis of Single-cell RNA Sequencing Data from the Mice Habenula has Identified a Variety of Distinct Neuronal Types.

The habenula is close to the midline and surrounded by the thalamus (Namboodiri et al., 2016) (Figure 3.5). It exhibits considerable conservation across vertebrate species. In mammals, it is anatomically and genetically subdivided into two main regions: medial habenula (mHb) and lateral habenula (lHb). mHb and lHb are accepted as homologous to dorsal and ventral habenula in zebrafish, respectively (Figure 3.5) (Hashikawa et al., 2020; Schmidt & Pasterkamp, 2017).

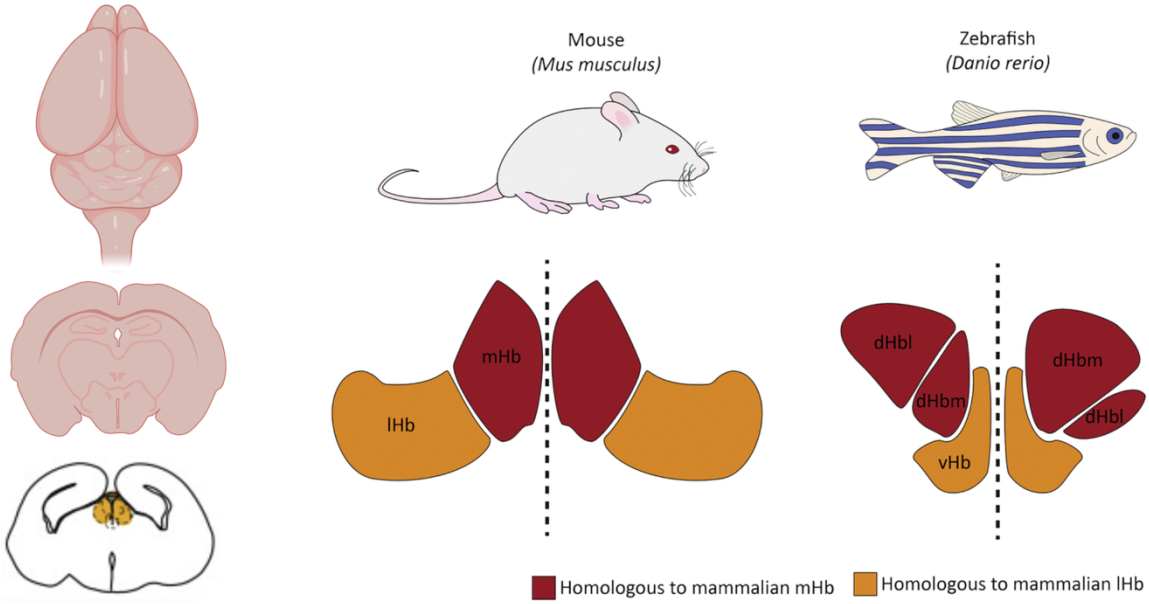

Figure 3.5: Illustration of comparative anatomy of the habenula in mice and zebrafish. (Left) Anatomical sections of the mouse brain show the habenula in different views: top view, coronal section, and sagittal section (van de Haar et al., 2022). (Middle) Anatomical illustration of adult mice habenula, divided into its subregions: medial habenula (mHb) and lateral habenula (lHb). (Right) Anatomical illustration of adult zebrafish habenula, divided into its subregions: the dorsal habenula lateral (dHbl), dorsal habenula medial (dHbm), and ventral habenula (vHb). The red color indicates regions homologous to the mammalian medial habenula (mHb), and the orange color indicates regions homologous to the mammalian lateral habenula (lHb). (Schmidt & Pasterkamp, 2017), (Created with BioRender).

The Allen Institute has recently compiled a comprehensive scRNA-seq dataset for the entire mice brain (Yao et al., 2023). This dataset includes annotations specifically for mHb and lHb glutamatergic cells, which are selected for use in this analysis. The mice habenula dataset from the Allen Institute comprises 6,053 cells and 32,285 genes derived from two animals (one female and one male). As the dataset had already been integrated, no further integration was performed (Figure 3.6c).

Before further analysis, I normalized and stabilized variance due to technical variability, such as differences in sequencing depth, which could lead to misleading conclusions about gene expression levels. Sctransform, which uses regularized negative binomial regression, is particularly effective for datasets exhibiting overdispersion, where variance exceeds the mean (Choudhary & Satija, 2022). To capture biological variation critical for identifying cellular heterogeneity, I selected highly variable genes using a method

implemented in Seurat. Based on this approach, 900 genes were identified as highly variable for further analysis (see methods) (Figure 3.6a). I chose the PC number 30 based on the standard deviation and heatmaps for each PC (Figure 3.6b).

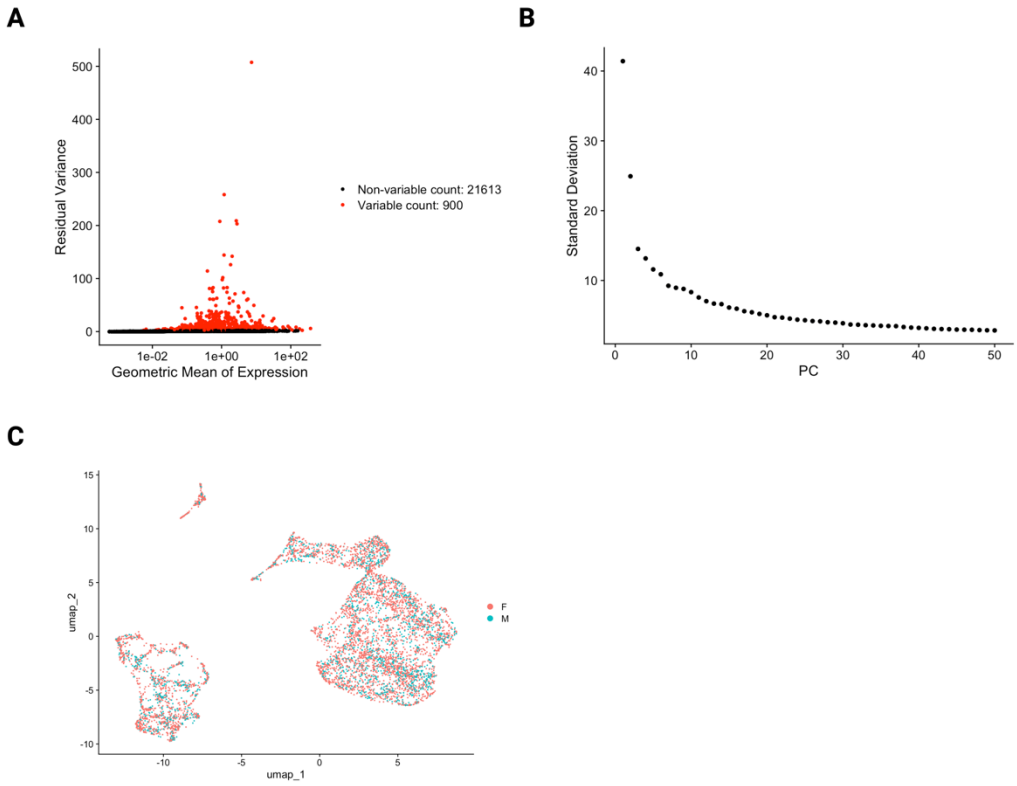

Figure 3.6: Analysis of the variable genes, PCA, and cell distribution of animals in the scRNA-seq mice data. (A) Scatter plot (Variable Feature Plot) depicting the residual variance as a function of each gene's geometric mean of expression. Genes are categorized into variable (red dots, 900 genes) and non-variable (black dots, 21613 genes) based on their variance. Variable genes show higher residual variance compared to non-variable genes. (B) Elbow plot showing the standard deviation of the principal components (PCs) derived from the PCA of the gene expression data. The first few PCs capture most of the variance, with a steep drop-off indicating the primary components of variation in the dataset. From this plot, the 30th principal component is selected for further analysis of the dataset. (C) UMAP plot illustrating the distribution of cells from two different animals integrated into the mice scRNA-seq dataset (female in red and male in cyan). The plot demonstrates the successful integration of cells from both animals.

To decide the resolution parameter, I used Clustree and constructed a tree diagram showing resulting clusters at different resolution values (Luke Zappia & Oshlack, 2018). By inspecting cluster stability and cell separation based on molecular markers, I chose a resolution value of 1.8 (Figure 3.7), which revealed 26 distinct cell groups within the mice

scRNA-seq dataset, which consists of habenular glutamatergic neurons. Cell removal was unnecessary since this dataset exclusively comprises neuronal cells and lacks non-neuronal cells.

Twenty-six transcriptionally distinct clusters are displayed in a UMAP (Figure 3.8a). Based on the molecular similarity of these clusters, I constructed a dendrogram for these clusters to delineate and comprehend the differences and similarities between cell groups. Additionally, a heatmap was created to show the most highly expressed genes in each cluster (Figure 3.8b). This heatmap provided detailed heterogeneous gene expression profiles across clusters. Clusters 1 and 2 show significant expression of genes such as *9330158H04Rik*, *chrm2*, *ntn1*, *D030068K23Rik*, *pmfbp1*, and *peg10*. Clusters 3 and 4 are characterized by the expression of *peg10*, *glra2*, *sst*, *sema3d*, and *lmo1*, with cluster 4 additionally expressing *pdgfd* and *cngb3*. Clusters 5 and 6 highlighted by high expression levels of *cngb3*, *gm49678*, *pde11a*, *cck*, *sned1*, and *calcrl*, with cluster 6 also expressing *wif1*, *rxfp2*, and *rspo4*. Clusters 7 to 10 share expressions of *sema3d*, *spon1*, *syt15*, *col8a2*, *rspo4*, and *lmo1*, with variations in specific gene expressions such as *tox*, *calcrl*, and *sned1*. Clusters 11 to 14 exhibit expressions of *spon1*, *syt15*, *col8a2*, *rspo4*, *tox*, *piezo2*, *trhde*, *vit*, and *sox2*. Clusters 15 to 20 features expressions of *rxfp2*, *syt15*, *lmo1*, *sned1*, *lamc3*, *calcrl*, and *piezo2*, with notable genes like *lsp1*, *igfbp7*, and *gulp1*. Clusters 21 to 26 are characterized by expressions of *9330158H04Rik*, *chrm2*, *ntn1*, *ll1rapl2*, *car2*, and *gulp1*. I also checked the *tac2* gene, which is used as a mHb marker, and *pcdh10*, which is an lHb marker (Hashikawa et al., 2020). *Tac2* is expressed in clusters 1, 2, and 21-26, while *pcdh10* is expressed in clusters 5-19. This analysis highlights the transcriptional diversity and complexity of the cellular clusters identified in mouse habenula.

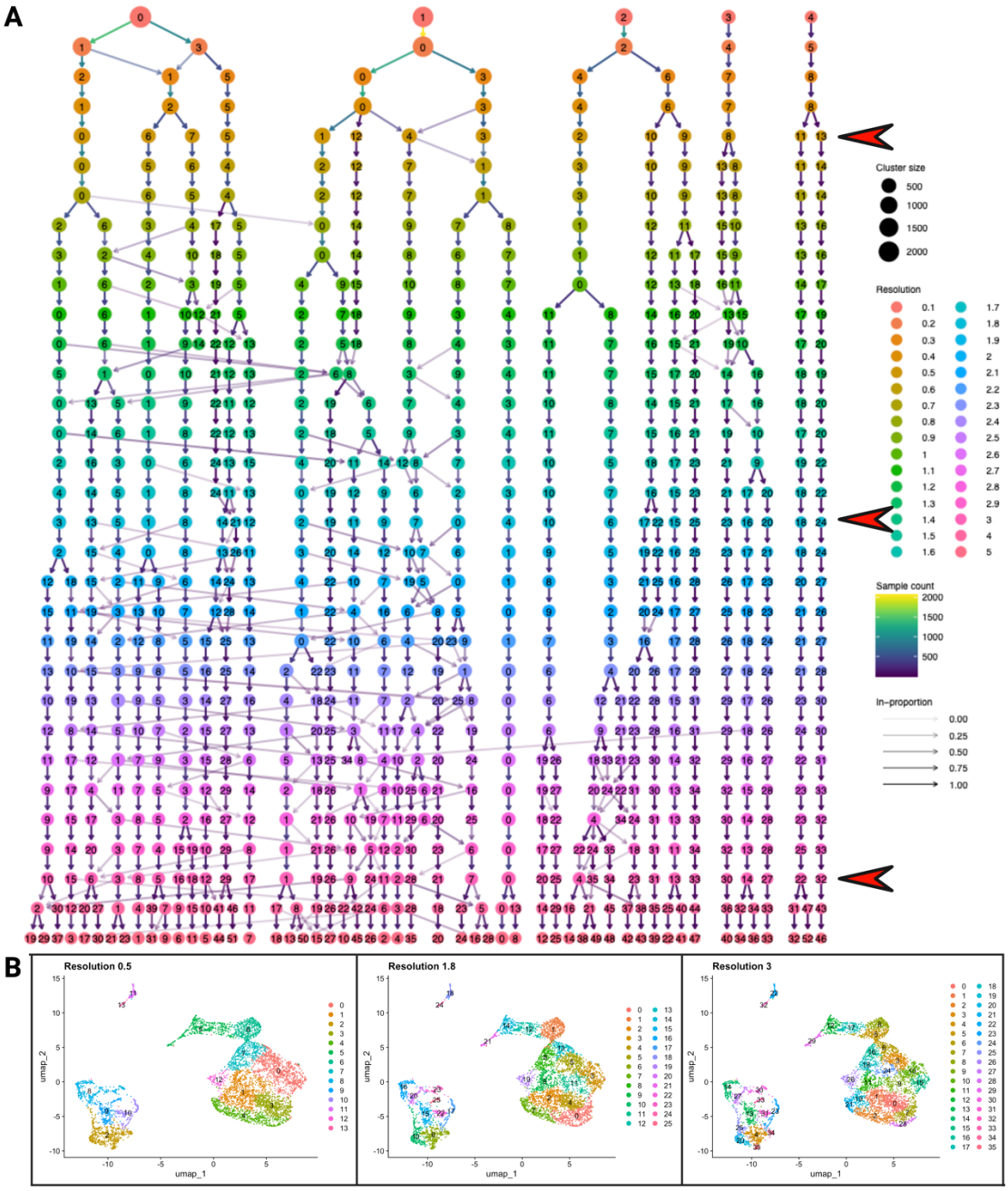

Figure 3.7: Cluster tree and UMAPs of the mice scRNA-seq data. (Top) Cluster tree for resolutions between 0.1 and 5; nodes are colored based on the value of resolution. Cluster size is indicated by the size of the nodes. Edges are colored to reflect the sample size, ranging from blue (fewer samples) to yellow (more samples). The transparency of the edges indicates the significance of the connection. Red arrows on the side indicate selected resolutions 0.5, 1.8, and 3 as examples. (Bottom) Each UMAP illustrates how the number and composition of clusters vary with different resolution settings (Created with BioRender).

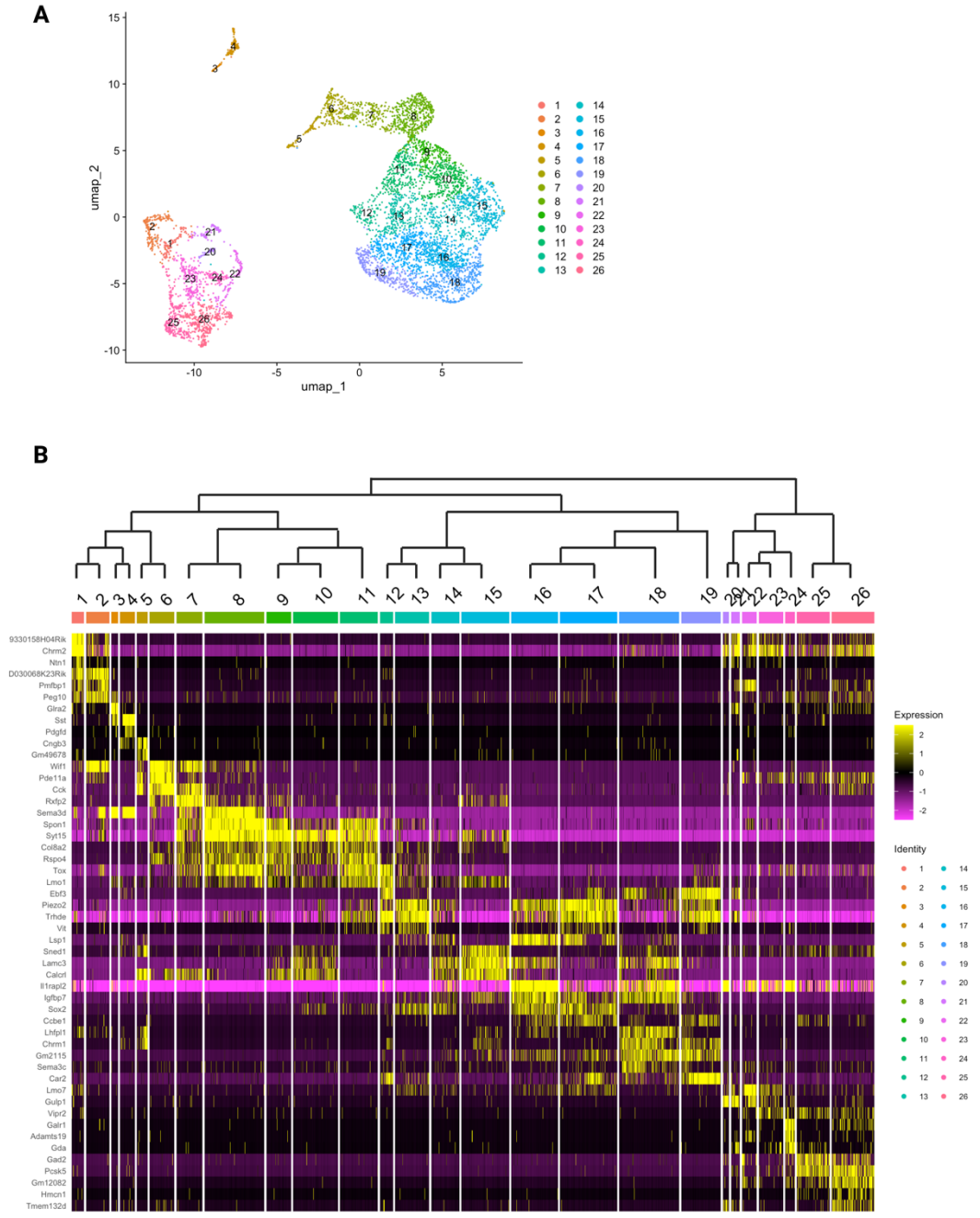

Figure 3.8: UMAP and the heatmap of the mice scRNA-seq data with dendrogram. (A) UMAP visualization of the mice scRNA-seq data, showing clusters sorted by molecular similarity. (B) Heatmap depicting gene expression profiles with an accompanying dendrogram illustrating the hierarchical clustering of the cells. Rows represent gene names, and columns represent clusters. The color intensity in the heat map indicates gene expression levels, with yellow indicating higher expression and black indicating no expression.

3.3 Transcriptional and Topographical Analysis Reveals Strong Spatial Organization of Cell Types in the Zebrafish Habenula.

Single-cell RNA sequencing (scRNA-seq) is significant in exploring the distinct cell types within the habenula and their molecular identities. However, this technology does not provide insights into the topology, spatial organization of molecularly similar cells, or their anatomical distributions on its own. To overcome these limitations, our lab utilized the Molecular Cartography technology developed by Resolve Biosciences (Benedict Nilges Resolve Biosciences) to generate the Atlas of Zebrafish Transcriptomic Encephalic Cytoarchitecture (AZTEC) dataset (Bredesen-Aa, 2024). This advanced technology enables sensitive detection and quantification of mRNA expression within tissue samples, enhancing our understanding of cellular architecture. For this project, I have given access to only the habenula in the AZTEC dataset (Figure 3.9). The dataset includes samples from two adult

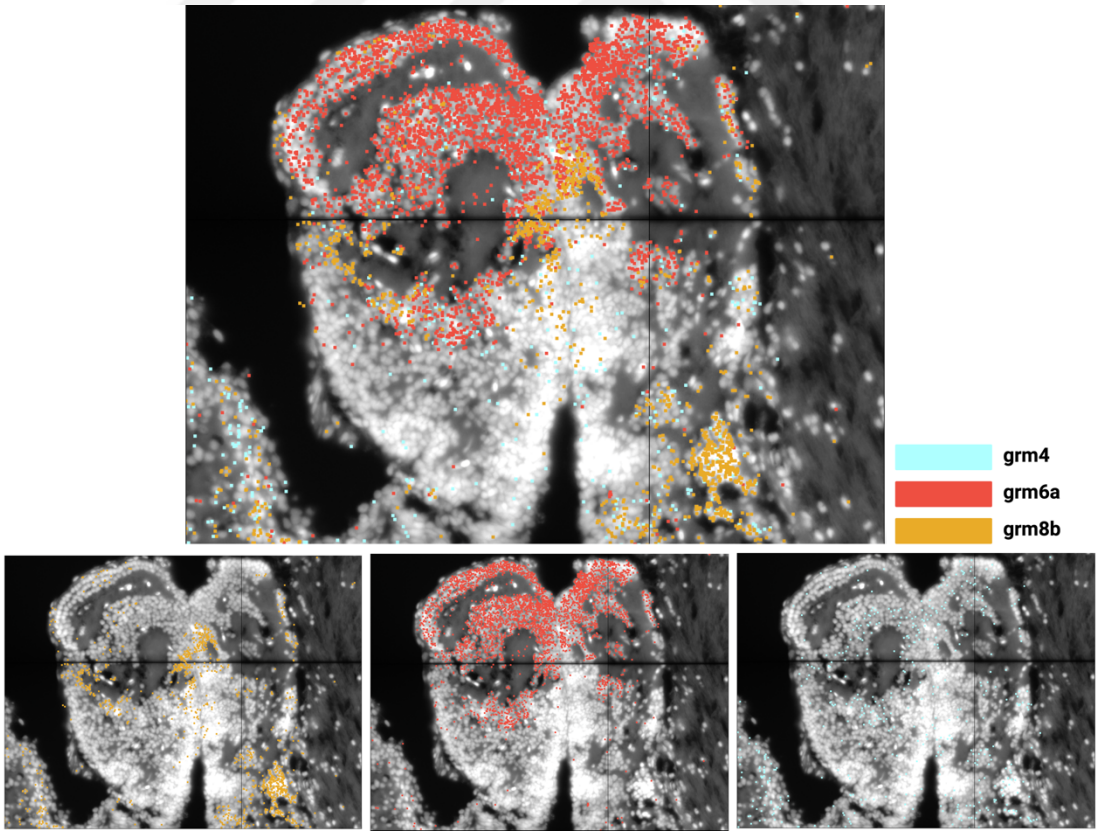

Figure 3.9: High-resolution molecular cartography on cryosection of the adult zebrafish habenula. (Top) Expressions of grm4 (blue), grm6a (red) and grm8b (orange). (Bottom, left) grm8b expression in the habenula, (bottom, middle) grm6a expression in the habenula and (bottom, right) grm4 expression in the habenula.

zebrafish habenula, each aged six months, encompassing 11,384 cells and 99 genes. In the original AZTEC dataset, samples from different animals were not integrated due to the lack of significant batch effects, enabling direct merging (Bredesen-Aa, 2024). To preserve the original data's consistency and given the absence of batch effects, I did not perform any additional integration for this AZTEC habenula dataset (Figure 3.10b). The details for constructing the AZTEC dataset and resolve technology are provided in the method section.

As discussed in the methods section, sctransform is employed to normalize and stabilize variance within the dataset. Using variable features plot implemented in the Seurat R package (methods), 60 out of 99 genes demonstrated significant variance in the habenula (Figure 3.10a). Contrary to the previous workflows, PCA is not utilized for this dataset due to the limited number of genes, which hampers the efficiency of PCA (Lenz et al., 2016).

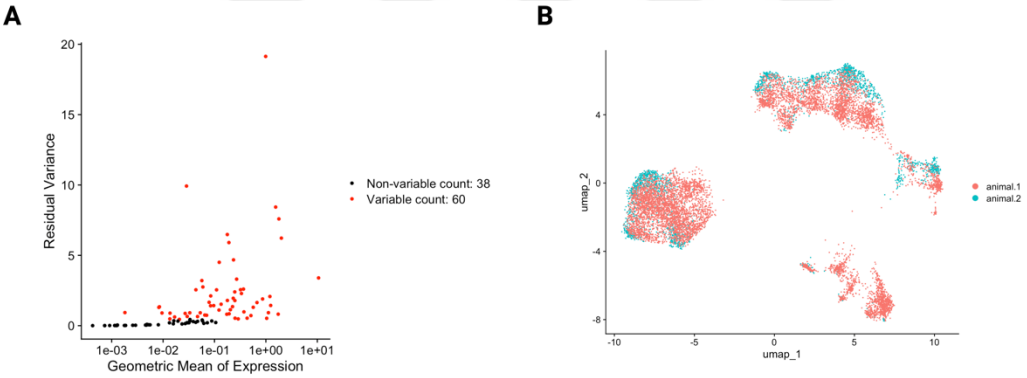

Figure 3.10: Analysis of the variable genes and cell distribution of two animals in the spatial zebrafish data. (A) Scatter plot (Variable Feature Plot) depicting the residual variance as a function of each gene's geometric mean of expression. Genes are categorized into variable (red dots, 60 genes) and non-variable (black dots, 38 genes) based on their variance. Variable genes show higher residual variance compared to non-variable genes. (B) UMAP plot illustrating the distribution of cells from two different animals in the AZTEC dataset (animal 1 in red and animal 2 in cyan). The plot demonstrates the successful merging of cells from both animals, indicating consistent spatial and transcriptional organization within the combined dataset.

To find the optimal resolution, I used clustree and UMAP for dimension reduction and analyzed the data across resolution values from 0.1 to 5 (Figure 3.11). Based on cluster stability and gene expression analysis, I found that a resolution of 1.5 is ideal for identifying distinct and stable cell groups in AZTEC habenula data.

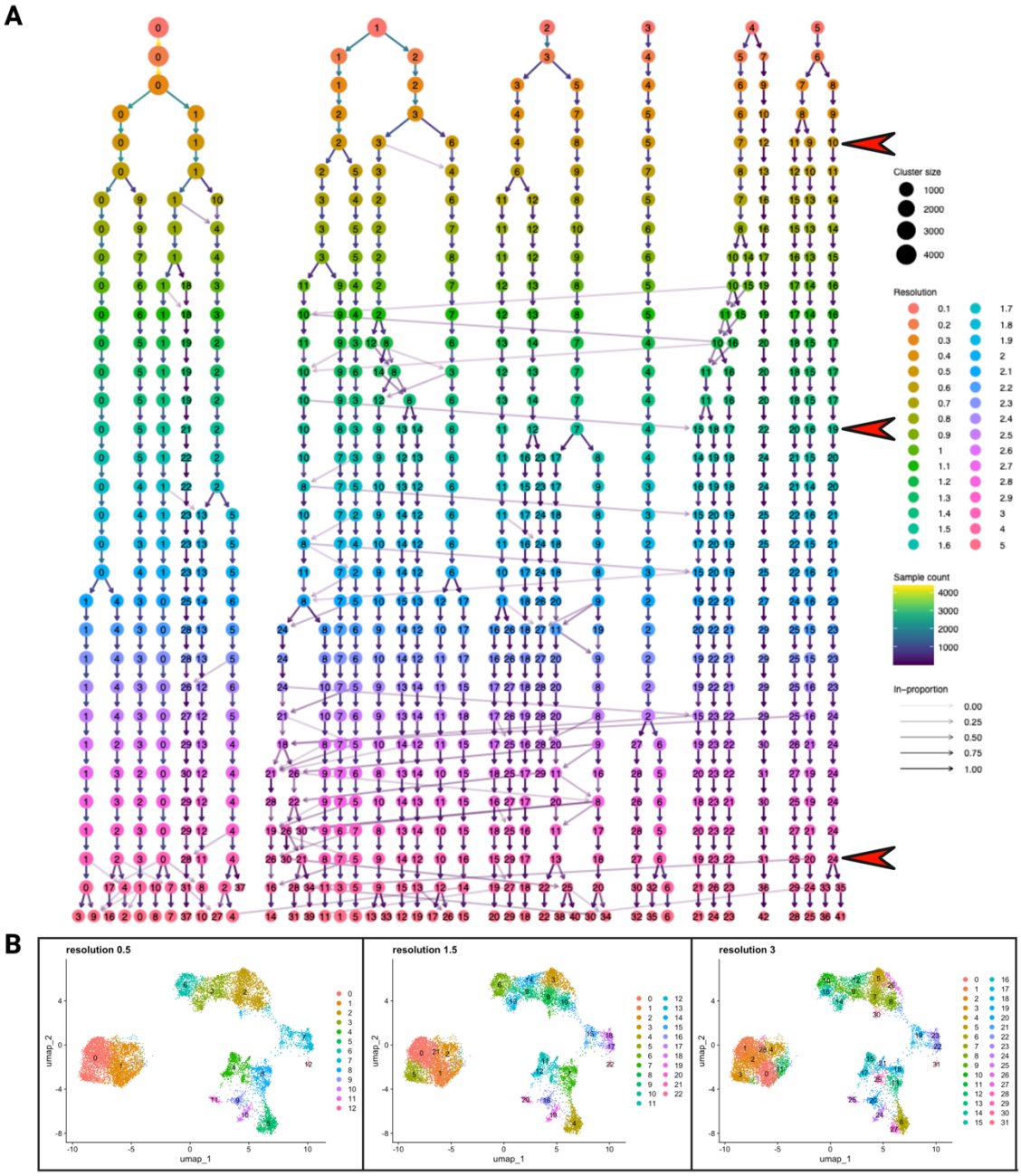

Figure 3.11: Cluster tree and UMAPs of the AZTEC habenula data. (Top) Cluster tree for resolutions between 0.1 and 5; nodes are colored based on the resolution value. Cluster size is indicated by the size of the nodes. Edges are colored to reflect the sample size, ranging from blue (fewer samples) to yellow (more samples). The transparency of the edges indicates the significance of the connection. Red arrows on the side indicate selected resolutions 0.5, 1.5, and 3 as examples. (Bottom) Each UMAP illustrates how the number and composition of clusters vary with different resolution settings (Created with BioRender).

This analysis identified 23 distinct cell groups, each with unique molecular identities. As previously noted, our previously used dataset lacks non-neuronal cell clusters. Consequently, I filtered out the non-neuronal cells from the AZTEC data. To achieve this, I first identified clusters that did not express the *snap25* gene (synaptosome-associated protein 25a), specifically clusters 7, 11, and 12. These clusters were found expressing glia-dominant genes such as *eaat2*, *gfap*, and *olig2* (Une et al., 2021; Zeis et al., 2015) and were excluded from further analysis. The analysis successfully delineated 20 transcriptionally distinct clusters, which can be visualized effectively using UMAP, as shown in Figure 3.12a. To further explore the molecular relationships and distinctions among these clusters, I constructed a dendrogram that illustrates their molecular similarities and differences for interpreting how these cell groups are related to one another (Figure 3.12b, top). Additionally, to provide a deeper insight into the molecular characteristics of each cluster, I generated a heatmap displaying the top expressed genes within each cluster (Figure 3.12b).

Based on this heatmap, gene expression profiles for each cell group can be identified. Notably, the *tacr3l* gene shows significant expression, specifically in clusters 1 to 5. The *pnoca* gene is expressed exclusively in clusters 1, 3, 6, 8, 17, 18, and 20. *Grm8b* is predominantly expressed in cluster 6, with additional expression observed in clusters 1 and 13. Similarly, *grm6a* is expressed in clusters 1 through 5, 9, 10, and 11, suggesting a broader distribution. The *pou3f1* gene exhibits its highest expression in cluster 6 and relatively higher expression in clusters 1, 4, 8, and 11.

The *glula* gene is expressed explicitly in clusters 13, 14, and 15. *Sst1.1* has strong expression in clusters 3 and 8 and comparatively higher expression in clusters 1 and 6, although lower than in clusters 3 and 8. *Slc17a6b* is expressed across clusters 3, 5, 6, 7, 8, 9, and 11. *Gng2* shows strong expression in clusters 4, 6, and 11, with some expression noted in clusters 5, 13, 14, and 15. The *scnlab1* gene is expressed in clusters 5, 7, 8, 9, 10, 11, 13, and 14, whereas *nptx2a* shows high expression in clusters 9, 10, and 11, with the highest in cluster 9.

Cluster-specific genes include *foxa1*, unique to cluster 6, and *sox1b*, specific to cluster 7. *Slc17a6a* shows high expression in clusters 7, 13, and 14, while *igf2a* is highly expressed in cluster 9. The *tac1* gene is strongly expressed in clusters 9 and 10 and relatively high in

cluster 14 compared to other clusters. *Chrna4b* shows high expression in clusters 9, 10, and 11, and *gad1a* is specific to cluster 12, which also exhibits high expression for *ppp1r1c* and *gabra1*.

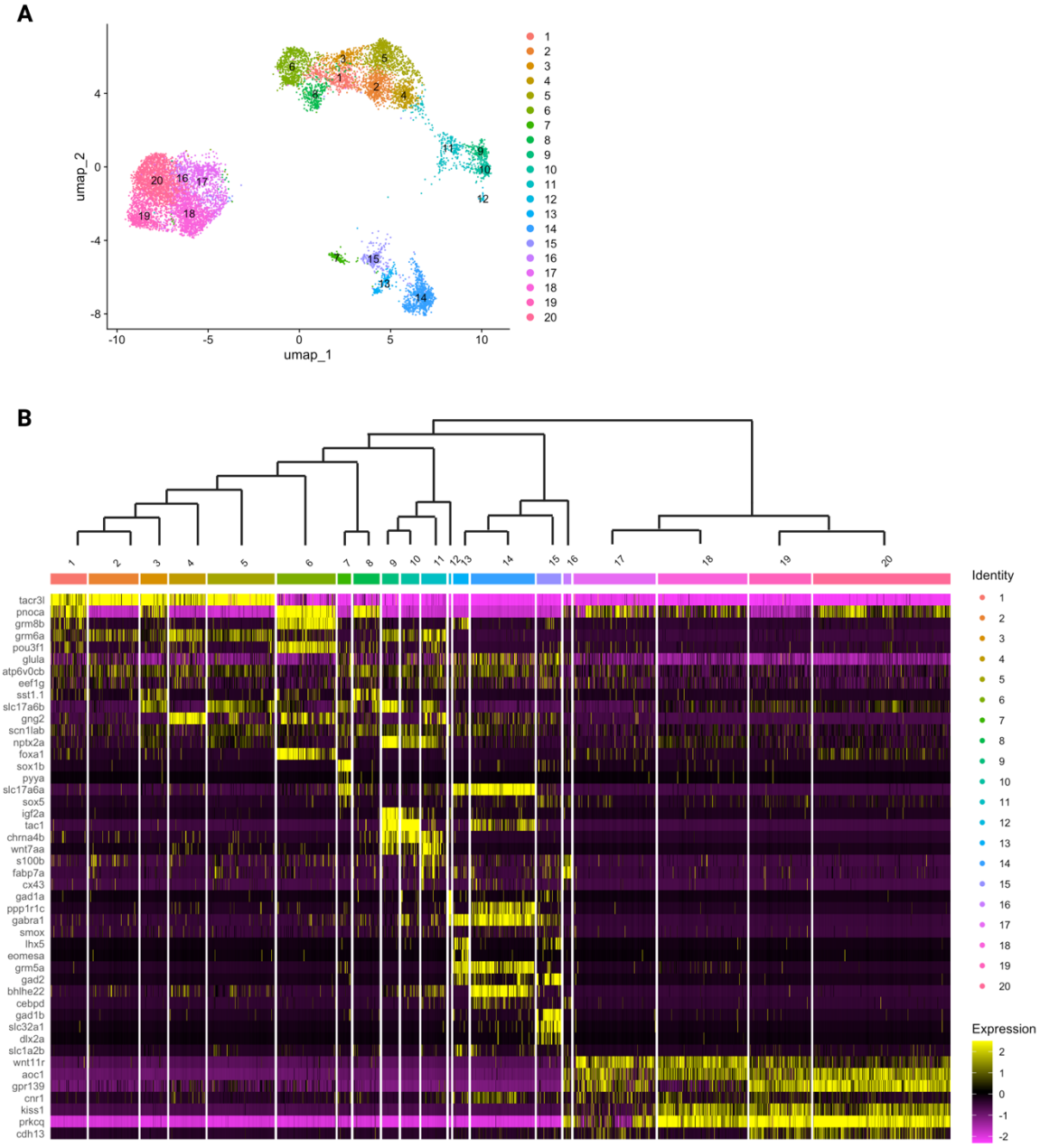

Figure 3.12: UMAP and the heatmap of AZTEC habenula data with dendrogram. (A) UMAP visualization of adult zebrafish AZTEC data, showing clusters sorted by molecular similarity. (B) Heatmap depicting gene expression profiles with an accompanying dendrogram illustrating the hierarchical clustering of the cells. Rows represent gene names, and columns represent clusters. The color intensity in the heat map indicates gene expression levels, with yellow indicating higher expression and black indicating no expression.

Clusters 13 and 15 are notable for expressing *wnt7aa*, *ppp1r1c*, *gabra1*, and *lhx5*. *Eomesa* and *grm5a* are specifically expressed in cluster 13, with *gad2* being highly expressed in clusters 13 and 15. *Gad1*, *slc32a1*, and *dlx2a* are expressed exclusively in cluster 15. The *fabp7a* gene is specific to cluster 16. Clusters 16 through 20 show expression for ventral markers *kiss1* and *aoc1*, suggesting these clusters might be in the ventral habenula (Pandey et al., 2018).

In the heatmap, clusters expressing the *aoc1* gene (potentially ventral clusters) also exhibit high expression levels for a broader set of genes, indicating a more homogeneous molecular identity within these clusters. Conversely, the non-*aoc1* expressing clusters (potentially dorsal clusters) display greater heterogeneity, with each cluster having distinct top expressed genes or sharing fewer highly expressed genes between clusters.

Following the clustering of cells based on molecular similarities, and identification of distinct cell groups within the zebrafish habenula, our next objective was to determine whether these cell groups also exhibit spatial organization. Spatial data includes the coordinates for each cell, making it possible to detect the anatomical positions of cells without being limited to marker genes.

To reveal the anatomical distribution of 20 clusters in AZTEC data, I generated the horizontal, coronal, and sagittal sections using the cell position coordinates (Figure 3.10) with the CoCy R package (Bredesen-Aa, 2024).

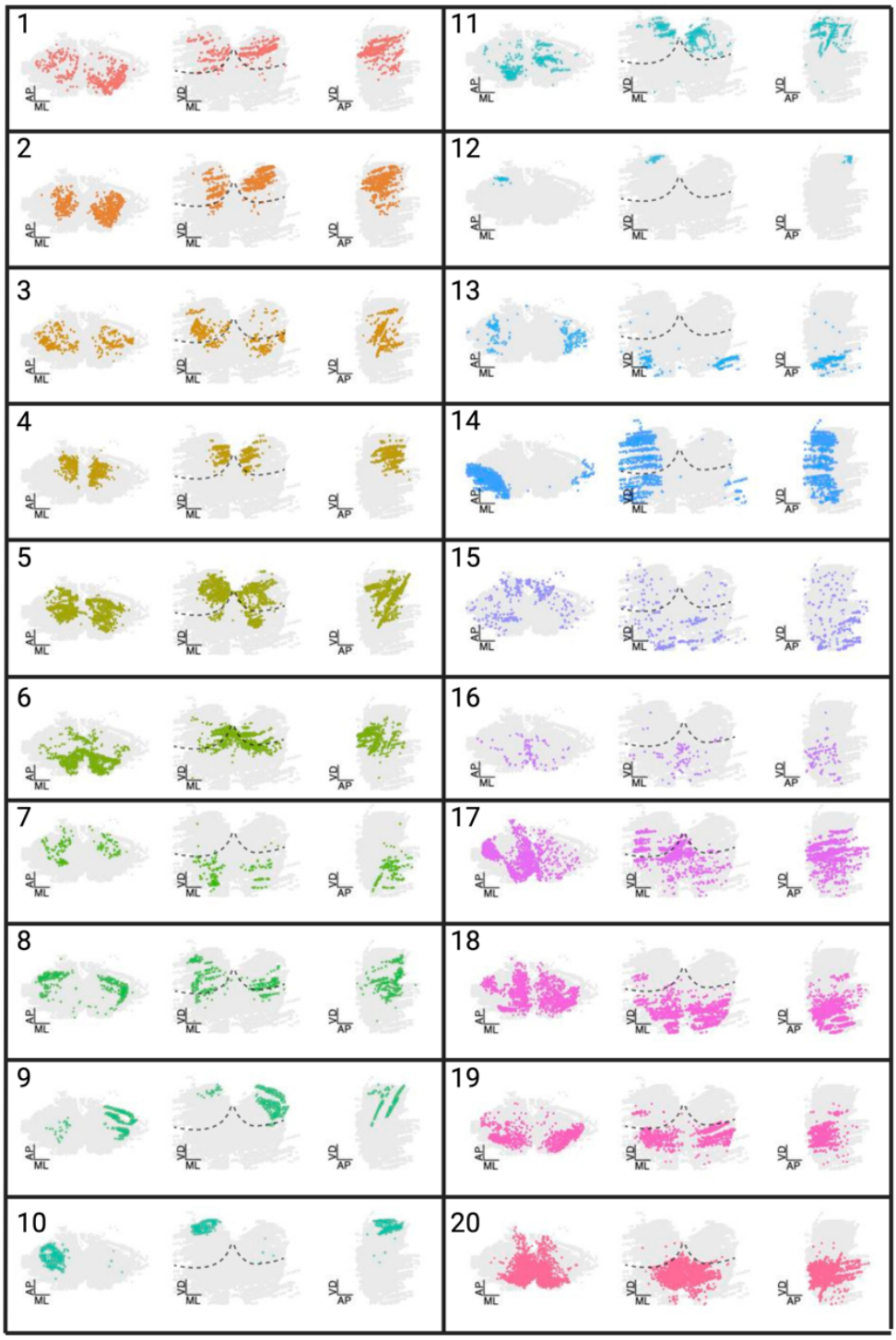

Figure 3.13: Spatial distributions of clusters in the AZTEC dataset. Horizontal (AP, ML), coronal (VD, ML), and sagittal (VD, AP) sections. The colors are chosen to match the cluster colors in UMAP. The numbers on the top right indicate the name of the cluster. The dashed line is constructed based on the expression of the *aoc1* gene, separating the dorsal (top) and ventral (bottom) domains of the habenula. AP: Anterior posterior axis, ML: Medial lateral axis, VD: Ventral dorsal axis. Scale bar indicating 50 μm. (Created with BioRender).

The anatomical distribution revealed that all cell groups (except cluster 15) show clear spatial organization (Figure 3.13). Notably, cluster 15 is distinguished as the only GABAergic neuronal cluster in our dataset due to robust expression of the GABA transporter gene *slc32a1* (Pandey et al., 2018). Clusters 1, 2, 4, 6, 8, 9, 10, 11, and 12 are localized in the dorsal part of the habenula. In contrast, clusters 7, 13, 16, and 20 exhibit a clear ventral anatomical distribution. The remaining clusters, specifically clusters 14, 15, 17, 18, and 19, do not show a definitive dorsal or ventral localization but are located/scattered in intermediate areas (Figure 3.13). However, it is important to note that the dashed line that separates the dorsal and ventral domains is created artificially based on the expression of the *aoc1* gene.

Horizontal, coronal, and sagittal sections clearly show the spatial organization of clusters. In order to quantify the spatial distribution, we examined the relationship between spatial position and topography by correlating RNA expression profiles of all pairs of neurons as a function of spatial distance between them (Figure 3.14). This plot illustrates how pairwise correlations change as the distance between neurons increases or decreases. The plot's high correlation at shorter distances indicates that molecularly similar cells are more likely to be located near each other. In contrast, lower correlations at greater distances suggest a spatial separation of different molecular identities. As a control (dashed line), I shuffled the spatial location of neurons in each cluster independent of the topographical organization. Those parts above the shuffled distributions show similarities above chance levels, while those below the shuffled distribution curve indicate similarities below chance levels.

In the plot it is seen that when distance is shorter (between 0 and 100 μm), high pairwise correlation is observed and it reduces as the distance increases suggesting that molecularly similar cells are more likely to be located near each other. At greater distances, lower correlation is observed, suggesting that as cells in zebrafish habenula become further apart, their molecular similarity also decreases (Figure 3.14). This further verifies the spatial organization within the habenula.

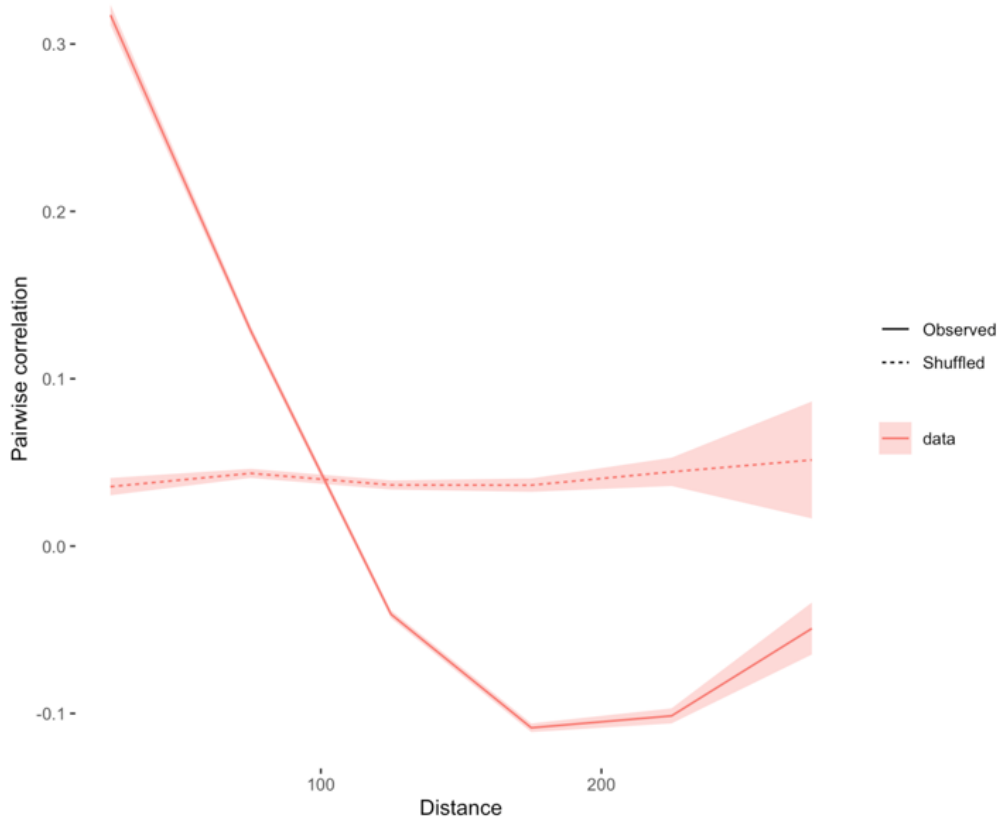

Figure 3.14: Correlation versus distance plot of the AZTEC habenula dataset. This plot shows the pairwise correlation between gene expression levels with respect to distance within the habenula. The Y axis is the pairwise correlation value, and the X axis is the distance in um. The solid line is the observed correlation, and the dashed line is the correlation for the randomized data (control). Distance is in um.

To gain deeper insights into the spatio-molecular organization of the Hb, I analyzed the dorsal and ventral clusters separately (Figure 3.15). Unique spatial correlation patterns specific to the dorsal and ventral clusters can be identified by examining these regions individually. In both plots, the observed pairwise correlation decreases as the distance increases as expected. However, the initial correlation in ventral clusters is higher than in the dorsal clusters, suggesting initially higher gene expression similarity among ventral cells. The decrease in correlation is more gradual compared to the dorsal clusters, indicating a slower decline in gene expression similarity with distance. Moreover, I also observed that vHb neurons exhibit higher correlations and, therefore, have lesser molecular diversity, whereas dHb neurons have lower correlations and are highly organized in space. Together,

these results suggested that dHb is more heterogeneous and exhibits stronger topography when compared to vHb.

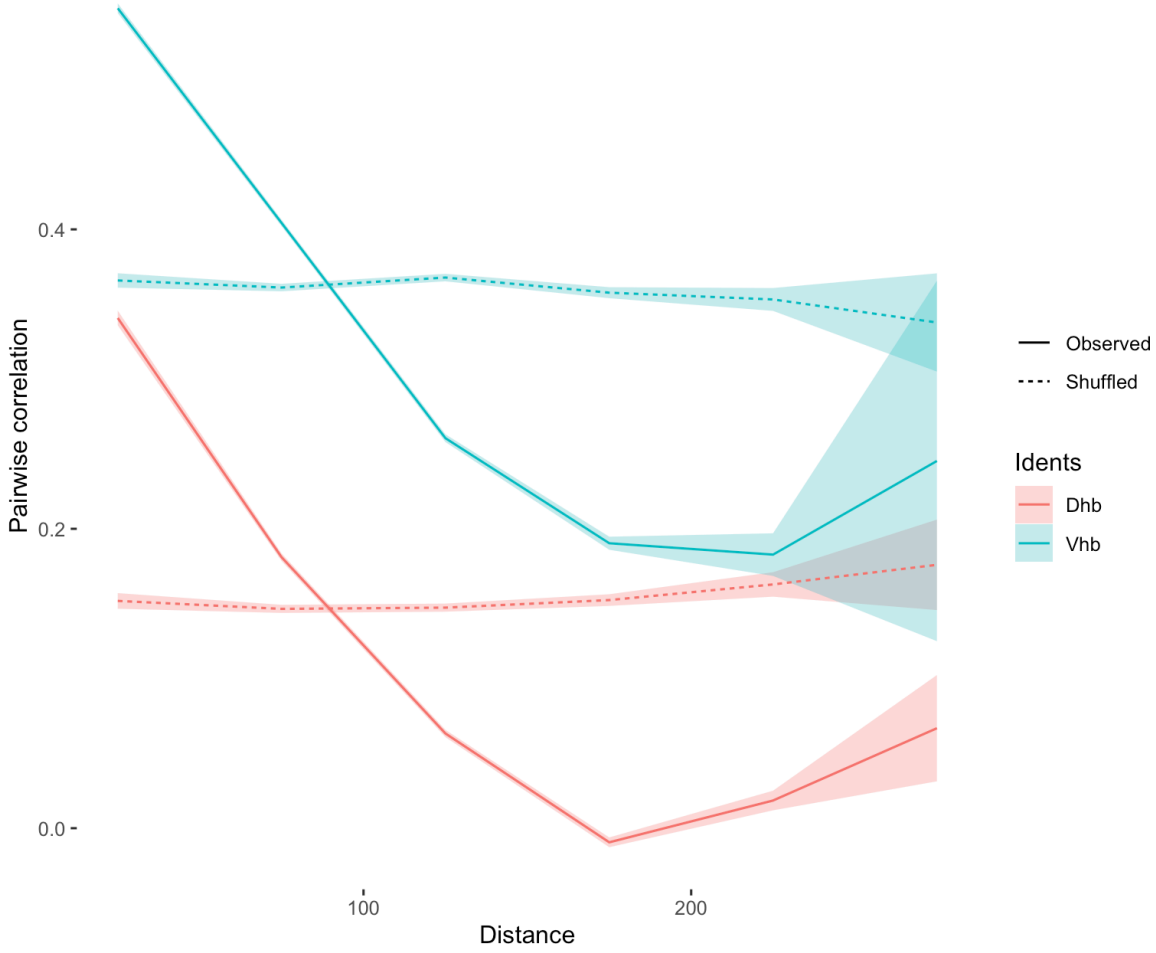

Figure 3.15: Correlation versus distance plots of dHb (red) and vHb (blue) in the AZTEC dataset. The y-axis is the pairwise correlation value; the x-axis is the distance. The solid line is the observed correlation, and the dashed line is the correlation for the randomized data (control). Distance is in μm.

In order to validate this spatial transcriptomic dataset, I integrated and compared the scRNA-seq adult zebrafish data with the spatial transcriptomics from the AZTEC adult zebrafish dataset (Figure 3.16). By comparing it to a published but smaller dataset, it can be confirmed that the gene expression patterns and identified cell types in AZTEC are consistent with what is known from the zebrafish scRNA-seq. Additionally, the AZTEC dataset provides critical spatial context and topography, an aspect lacking in traditional scRNA-seq. This comparative analysis allows for determining which genes are expressed in which clusters and identifying their specific locations within the habenula. Despite the limited gene

pool in the AZTEC dataset, integrating and comparing it with the scRNA-seq data offers more profound insights into the gene expression patterns of the clusters that might otherwise remain undetected in the AZTEC dataset alone.

In order to integrate, I utilized from the CoCy R package (Bredesen-Aa, 2024). First, the EggNOG map is created to obtain orthologous groups of genes for mice and zebrafish (Muller et al., 2010). CoCy integrates datasets based on Seurat's anchor-based Canonical Correlation Analysis (CCA) and calculates similarity among clusters with the K Mutual Ratio (KMR) across their joint embedding (see methods). This tool created a comparison matrix to evaluate the gene expression similarities and cluster overlaps between the AZTEC dataset and the scRNA-seq data from adult zebrafish (Figure 3.16).

In this comparison matrix, as expected, the scRNA-seq clusters marked with the ventral habenula marker kiss1 strongly correlate with the ventral clusters in AZTEC (Figure 3.16). This trend is also observed for the dorsal clusters, which correlate well with their respective counterparts in AZTEC. In the comparison matrix, several AZTEC clusters strongly correlate with specific clusters in the adult zebrafish single-cell RNA-seq dataset, indicated by intense and warmer KMR color overlaps. For example, dHb cluster 10 and 9 from AZTEC strongly correlate with dHb cluster 13 from the single-cell data, and dHb cluster 12 from AZTEC aligns closely with dHb cluster 3. Between the ventral cells of both dataset, overlap profile is more scattered compared to dorsal clusters. Additionally, the clusters (1 and 2) that remained unknown in terms of their anatomical position due to lack of marker gene expression, show overlap with the vHb cell groups of spatial zebrafish dataset, indicating these clusters might be in ventral habenula region.

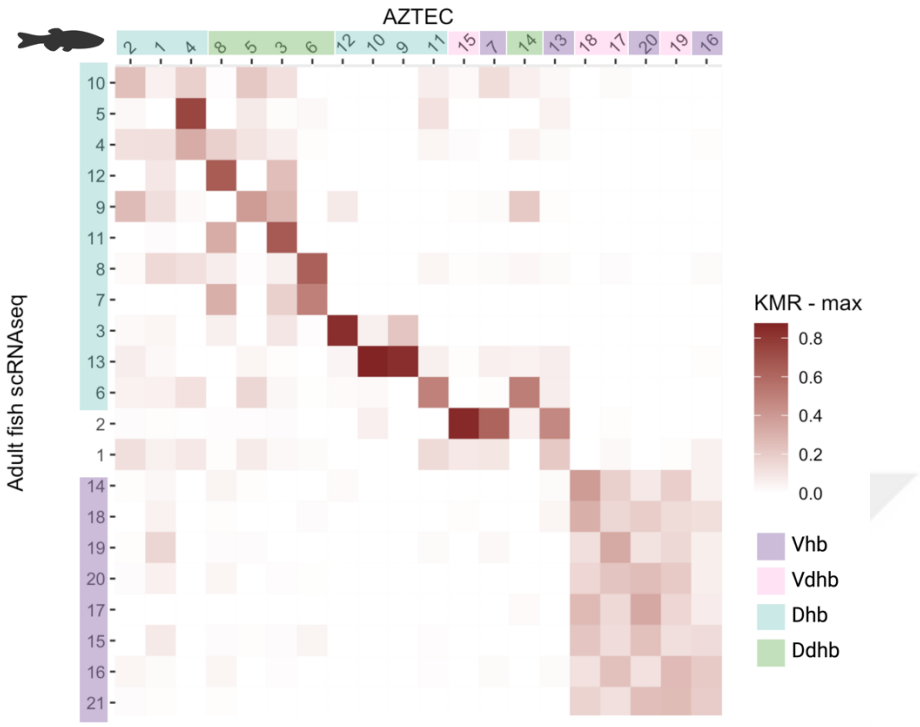

Figure 3.16: Comparison matrix of the AZTEC and zebrafish scRNA-seq data. Columns are clusters of AZTEC data; rows are clusters of adult zebrafish scRNA-seq data. The intensity of color shows a higher KMR value. Purple color indicates ventral habenula cluster, pink indicates ventral dominated habenula cluster (meaning ventral cluster with some cells in the dorsal domain), blue indicates dorsal habenula cluster; green indicates dorsal dominated habenula cluster (meaning dorsal cluster with some cells in the ventral domain). KMR: K-Mutual ratio.

To further verify the reliability of the anatomical distribution of gene expression patterns, I performed the hybridization chain reaction (HCR) combined with a tissue-clearing protocol (see Methods). I selected two genes that our lab is interested in, and we know that they are predominantly expressed in the dorsal and ventral domains of zebrafish habenula: *grm6a* and *aoc1*, respectively. Using this technique, I visualized the expression of these genes in the whole brain of 7-month-old adult zebrafish, focusing on the habenula (Figure 3.17). As shown in the figure, the gene expression profiles of *aoc1* and *grm6a* closely match the AZTEC spatial dataset and the HCR results. This alignment further verifies the reliability and accuracy of the AZTEC dataset.

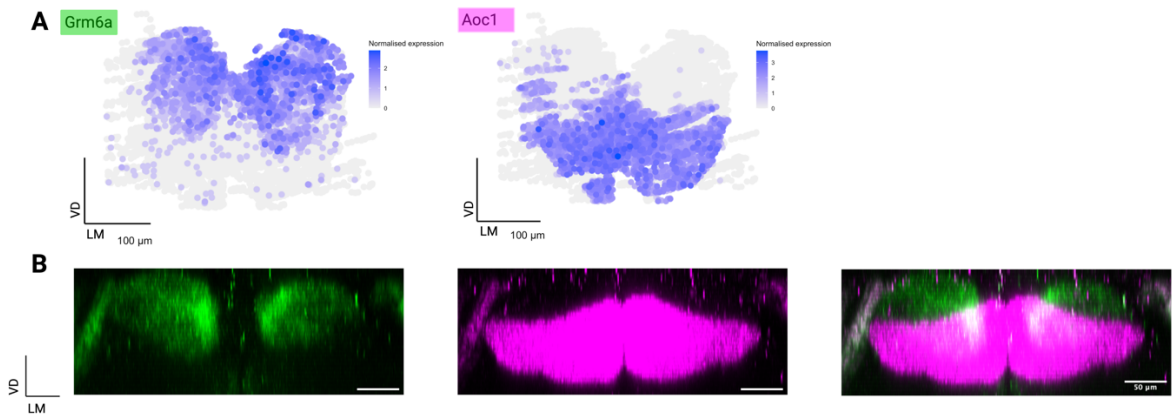

Figure 3.17: Comparison of *grm6a* and *aoc1* gene expressions in the AZTEC dataset and HCR results. (A) Coronal section (VD, ML) of gene expression plot for (left, green) *grm6a* and (right, pink) *aoc1*. The bar indicates the normalized expression (from blue to gray, highest to lowest, respectively.) The scale bar is 100 µm. (B) 7-month adult zebrafish, coronal view (VD, LM) of confocal images of (left, green) *grm6a*, (middle, pink) *aoc1*, and (right) both *grm6a* and *aoc1*. The scale bar is 50 µm. VD: Ventral, dorsal. LM: Lateral, medial. (Created with BioRender).

3.4 Transcriptional and Topographical Analysis Reveals Strong Spatial Organization of Cell Types and Microcircuits in the Mouse Habenula

Single cell RNA-seq data from mice allowed us to categorize cell types and molecular identities and predict the anatomical distribution of these cells based on the marker genes. The next step involves analyzing the spatial transcriptomic dataset for the mouse habenula. Recently, a spatial transcriptomic dataset was constructed using multiplexed error-robust fluorescence in situ hybridization (MERFISH) (Zhang et al., 2023). This dataset is available for visualization in the "Allen Brain Cell Atlas". For this project, I utilized the subset of the MERFISH dataset labeled as "habenular," comprising 5759 cells and 500 genes (Figure 3.18). The cells within this dataset are annotated as glutamatergic habenular neurons, meaning that the data exclusively comprises this subtype of neurons in the habenula. Additionally, no integration is done since this dataset only includes one male mouse.

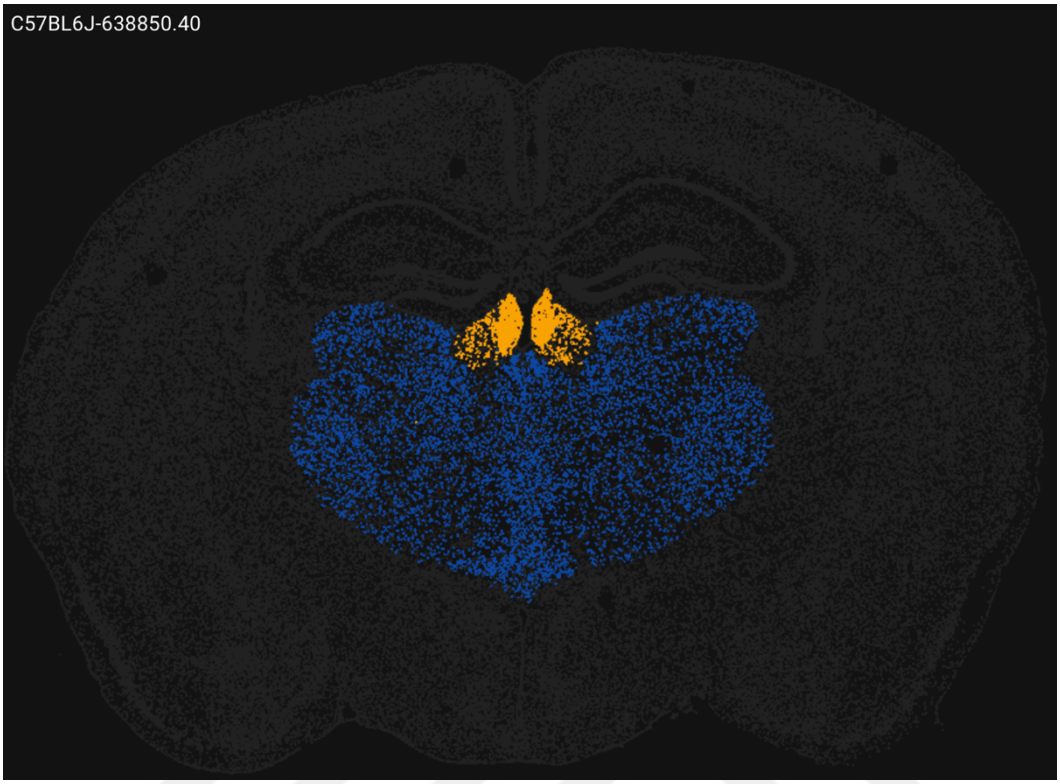

Figure 3.18: Coronal section of the MERFISH dataset showing habenula (yellow) and thalamus (blue). (Section number: C57BL6J-638850, image is from Allen Brain Atlas).

Similar to the previous data analysis in this project, I normalized and stabilized variance with sctransform, a method well-suited for datasets with high variance. Following normalization, I used a Seurat tool to identify 300 genes with significant variability (see methods) (Figure 3.19). For principal component analysis, I selected 26 principal components based on their variance in the elbow plot (Figure 3.19).

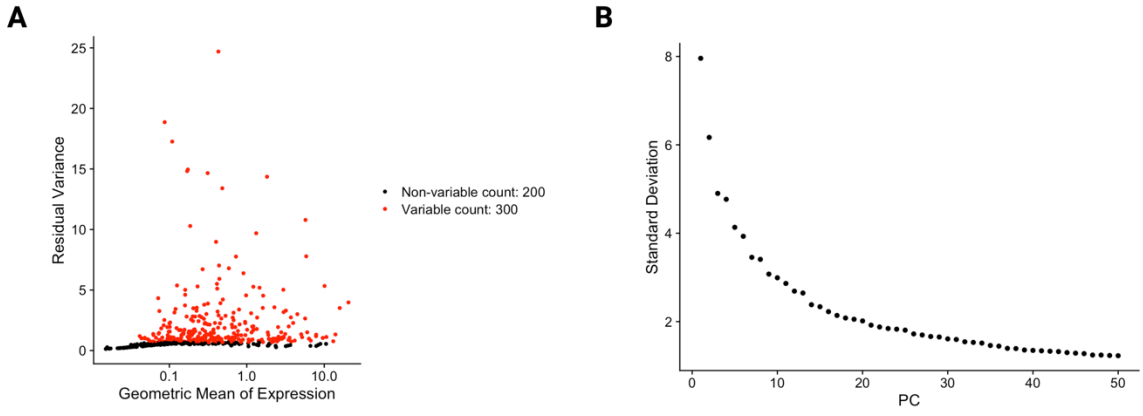

Figure 3.19: Analysis of the variable genes and PCA in the spatial mouse data. (A) Scatter plot (Variable Feature Plot) depicting the residual variance as a function of each gene's

geometric mean of expression. Genes are categorized into variable (red dots, 300 genes) and non-variable (black dots, 200 genes) based on their variance. Variable genes show higher residual variance compared to non-variable genes. (B) Elbow plot showing the standard deviation of the principal components (PCs) derived from the PCA of the gene expression data. The first few PCs capture most of the variance, with a steep drop-off indicating the primary components of variation in the dataset. From this plot, 26 is selected as the principal component for the MERFISH dataset.

To optimally cluster the cells, I used Clustree, testing different resolutions ranging from 0.1 to 5. The resolution of 1.7 was the most suitable for further analysis based on the cluster stability and separation (Figure 3.20).

This analysis revealed 28 distinct cell groups within the mouse spatial dataset, which consist of only habenular glutamatergic neurons. Since this dataset exclusively comprises neuronal cells, removing non-neuronal cells was unnecessary. These 28 unique transcriptional clusters are sorted based on their molecular similarity and visualized in the UMAP (Figure 3.21a). I constructed a dendrogram based on their molecular similarities to further explore the variations and connections among these cell groups. I generated a heatmap highlighting the genes most prominently expressed in each cluster (Figure 3.21b).

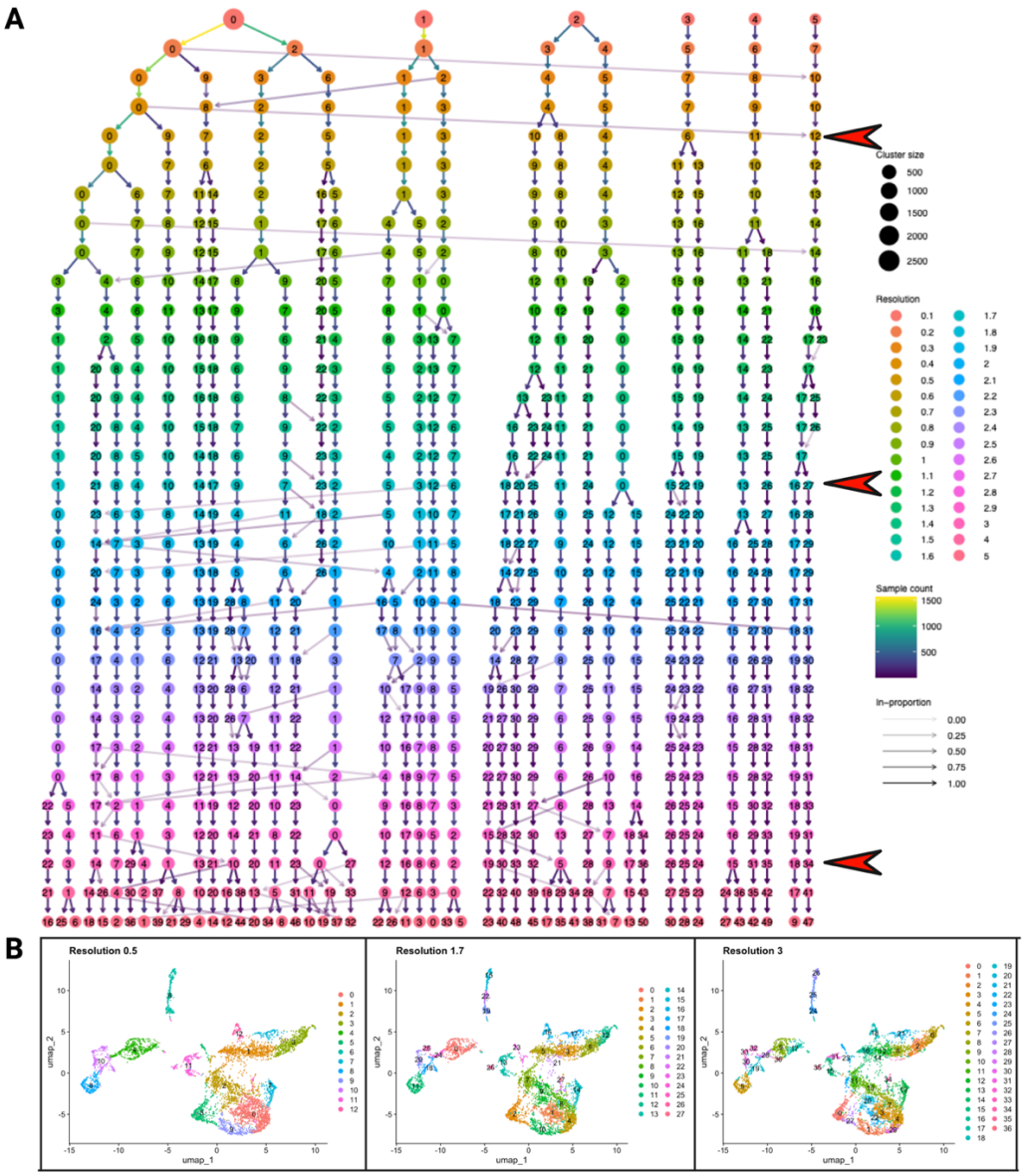

Figure 3.20: Cluster tree and UMAPs of the mouse MERFISH habenula data. (Top) Cluster tree for resolutions between 0.1 and 5; nodes are colored based on the resolution value. Cluster size is indicated by the size of the nodes. Edges are colored to reflect the sample size, ranging from blue (fewer samples) to yellow (more samples). The transparency of the edges indicates the significance of the connection. Red arrows on the side indicate selected resolutions 0.5, 1.7, and 3 as examples. (Bottom) Each UMAP illustrates how the number and composition of clusters vary with different resolution settings (Created with BioRender).

Cluster 1 showed strong expressions of genes such as *slc7a11*, *aqp4*, and *gja1*. Cluster 2 similarly expressed *slc7a11*, *aqp4*, *gja1* and *adamts9*, but differed by expressing *chrnb3*, *gpr4*, *fgf10*, *lsp1*, *synpr* and *slc5a7* instead. Cluster 3 was characterized by expressions of *C030029H02Rik*, *opalin*, *mog*, and *doc5*, while Cluster 4 showed expressions of *fras1*, *scn7a*,

rab3b, *fam163a* and *unc13c*. Cluster 5 expressed *c1ql1*, *spon1*, *ntn1*, *fbn2*, and *vcan*. Cluster 6 showed expressions of *cldn5*, *igf2*, and *abcc9*. Cluster 7 expressed genes such as *cnr1*, *kcng1*, *rprm* and *adcyap1*, which were also seen in cluster 8 along with *wif1*, *calb1*, *pde11a*, *dock5*, and *syt10*. Cluster 9 exhibited expressions of *fras1*, *scn7a*, *rab3b*, *cnr1*, *fam163a*, *unc13c*, *pde11a*, and *grm8*. Cluster 10 expressed genes like *cnr1*, *kcng1*, *adcyap1*, *wif1*, *fam163a*, *unc13c*, *pde11a*, *grm8*, *dock5* and *pde11a*. Cluster 11 showed *cnr1*, *pde11a*, *dock5*, *pou6f2*, and *tfap2d* expressions. Cluster 12 expressed *ecel1*, *ltbp1*, *vwa5b1*, *cartpt*, *th*, *syt10*, *slit3* and *trpc7*. Cluster 13 included expressions of *calb1*, *ecel1*, *ltbp1*, *vwa5b1*, *glra2*, *slit3* and *nxph4*. Cluster 14 showed expressions of *ltbp1*, *th*, *nxph4*, *cartpt*, *lsp1*, *syt10*, and *trpc7*. Cluster 15 expressed *chrnb3*, *gpr4*, *syndig1l*, *fgf10*, *synpr* and *slc5a7*, while Cluster 16 showed *chrnb3*, *gpr4*, *lsp1*, *synpr*, *slc5a7*, and *trhde*. Cluster 17 exhibited expressions of *chrnb3*, *gpr4*, *syndig1l*, *fgf10*, *synpr* and *slc5a7*. Cluster 18 showed *chmb3*, *synpr*, *ebf3*, *syt2* and *trhde*. Cluster 19 included *pou6f2*, *tfap2d*, *lhx8*, *chrnb3* and *synpr*. Cluster 20 expressed *arhgap25* and *ctss*. Cluster 21 showed expressions of *th*, *gpr4*, *syt10*, *trpc7*, *spon1*, and *rprm*. Cluster 22 included *calb1*, *th*, *nxph4*, *syt10*, *spon1* and *rprm*. Cluster 23 expressed *trhde*, *dgkk*, *lncenc1*, *gabrq*, *popdc3*, *ret* and *chrm2*. Cluster 24 showed expressions of *nrg1*, *fras1*, *scn7a*, *gabrq* and *gad2*. Cluster 25 included *vcan*, *kcng1*, *syt2*, *kcns3*, *scn4b*, *pvalb*, *popdc3*, *drd2*, *chrmn2* and *parm1*. Cluster 26 expressed *calb1*, *syt2*, *arhgap36*, *popdc3*, *ret*, *parm1* and *chrm2*. Cluster 27 included expressions of *kcng1*, *drd2*, *popdc3*, *ret*, *parm1* and *chrm2*. Finally, cluster 28 showed expressions of *popdc3*, *stxbp6*, *kcnab3*, *parm1*, and *chrm2*.

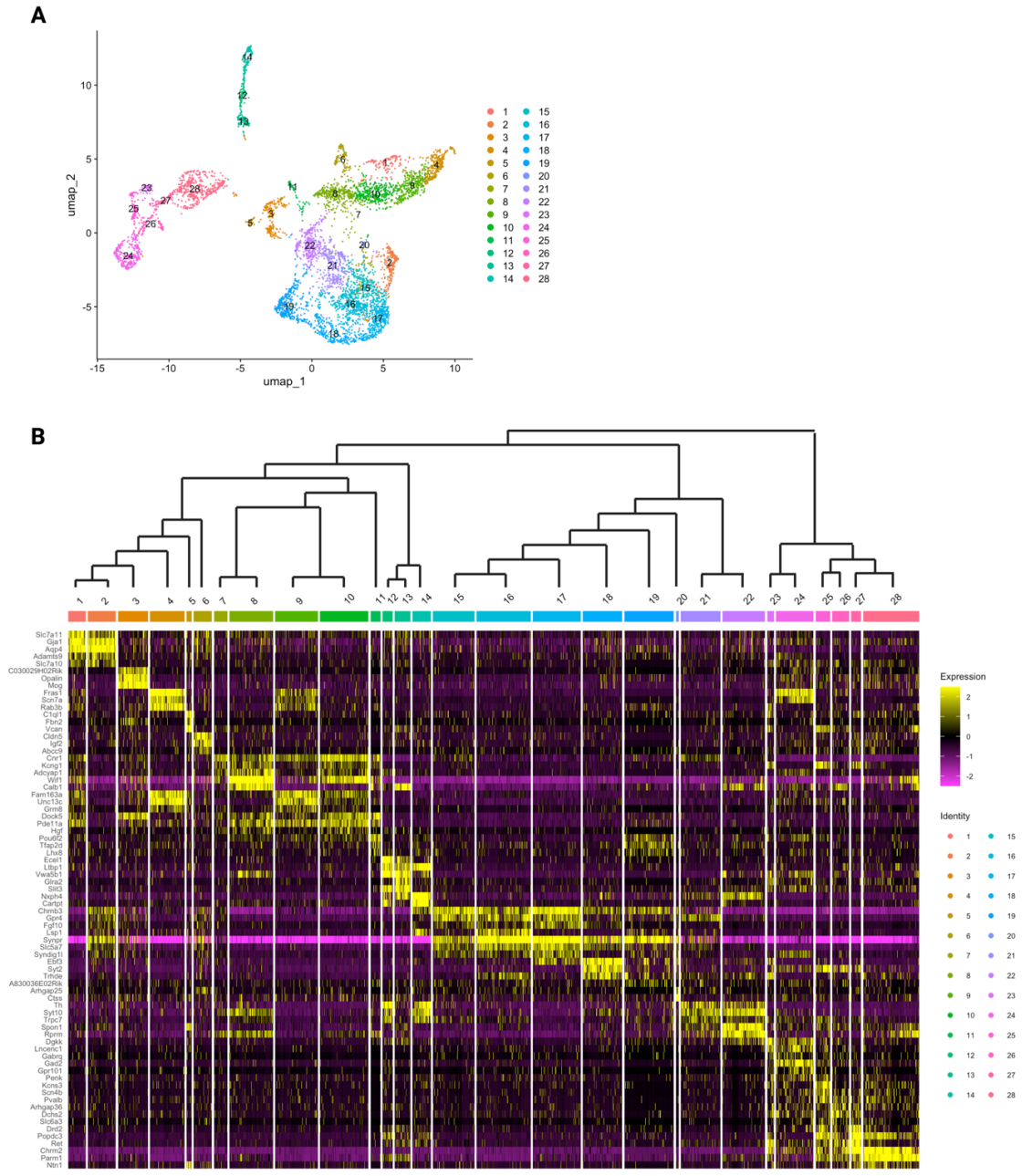

Figure 3.21: UMAP and the heatmap of the mouse MERFISH habenula data with dendrogram. (A) UMAP visualization of adult mouse MERFISH data, showing clusters sorted by molecular similarity. (B) Heatmap depicting gene expression profiles with an accompanying dendrogram illustrating the hierarchical clustering of the cells. Rows represent gene names, and columns represent clusters. The color intensity in the heat map indicates gene expression levels, with yellow indicating higher expression and black indicating no expression.

Based on *tac2* expression, clusters 1-22 are predicted to be mHb clusters, and clusters 24-28 are predicted to be lHb clusters. Similar to scRNA-seq mice, zebrafish, and AZTEC results, the gene expressions exhibit a scattered distribution profile in lHb clusters. On the other hand, mHb clusters exhibit unique and specific gene expression profiles, suggesting a greater molecular heterogeneity than lHb clusters.

After clustering cells based on molecular similarities, I aimed to explore whether these cell groups also show spatial organization. The advantage of spatial data is that it allows us to examine the anatomical positions of cell groups directly without being restricted to marker genes. To achieve this, I created spatial images of the MERFISH dataset using each cell's ccf coordinates (see methods) with CoCy, as illustrated in Figure 3.22. Horizontal, coronal, and sagittal views in the figure clearly demonstrate the spatial organization of clusters in the mouse habenula, similar to zebrafish habenula. Clusters 1-22 are primarily distributed in the mHb, and clusters 24-28 are predominantly positioned in the lHb. It is important to mention that the dashed line is created arbitrarily based on the coronal section in the Allen Institute portal.

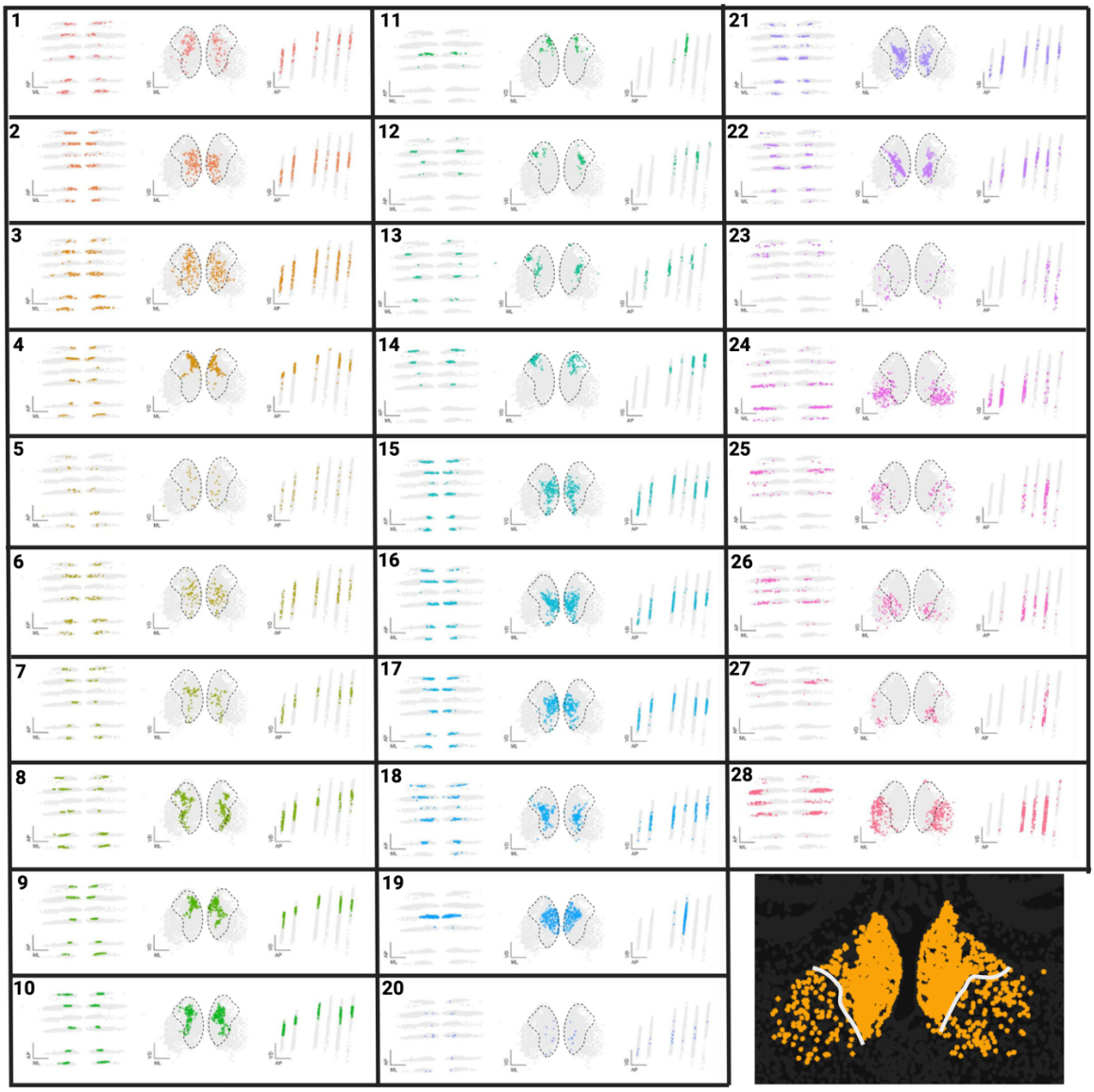

Figure 3.22: Spatial distributions of clusters in the MERFISH dataset. Horizontal (AP, ML), coronal (VD, ML), and sagittal (VD, AP) sections. The colors are chosen to match the cluster colors in UMAP. Numbers on the top right indicate the name of the cluster. The dashed line created based on the coronal section in Allen institute portal (right, bottom), separating the medial and lateral domains of the habenula. AP: Anterior posterior axis, ML: Medial lateral axis, VD: Ventral dorsal axis. Scale bar indicating 250 μm. (Created with BioRender).

The next objective is quantifying this topographical distribution of distinct cell groups in the mouse habenula (Figure 3.23). For this purpose, the relationship between spatial position and topography is studied by using a correlation versus distance plot. This plot shows how pairwise correlations differ with spatial distance. In the plot, a high correlation at shorter distances indicates that molecularly similar cells tend to be located near each other. In comparison, lower correlations at greater distances suggest a spatial separation of different molecular profiles. It is seen that pairwise correlation is higher (peaking around 0.3) at shorter distances (0-200 µm), and as the distance increases, the pairwise correlation decreases and falls to near 0 around 500 µm. This indicates that anatomically closer Hb neurons exhibit higher molecular similarity.

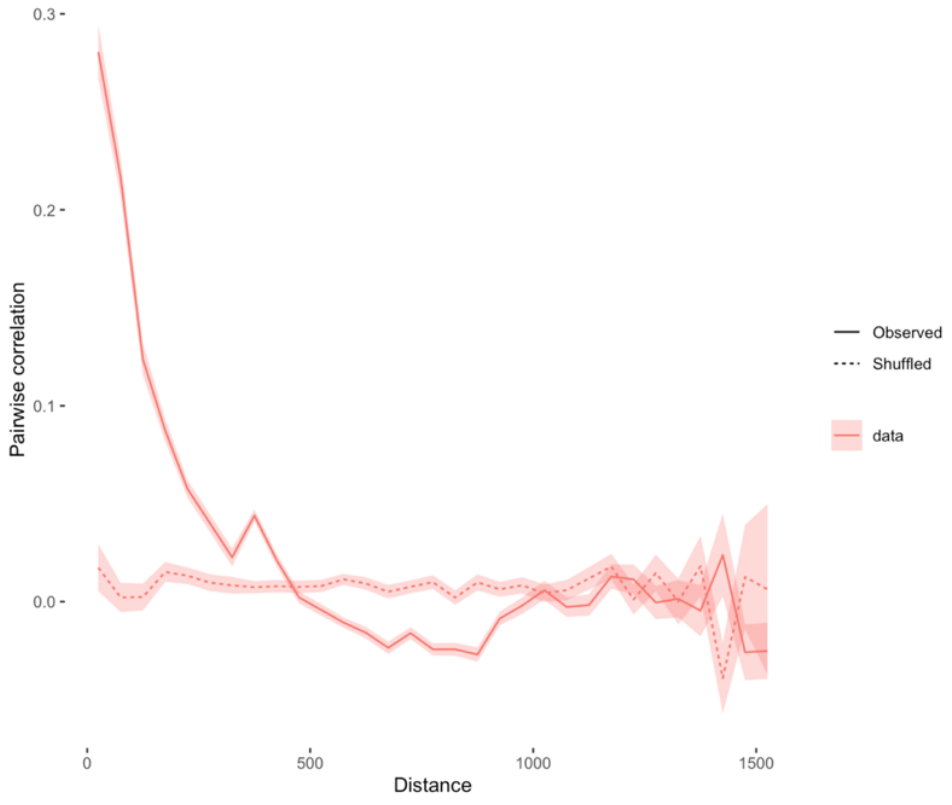

Figure 3.23: Correlation versus distance plot of the habenula in the MERFISH dataset. This plot shows the pairwise correlation between gene expression levels with respect to distance within the habenula. The y-axis is the pairwise correlation value; the x-axis is the distance in um. The solid line is the observed correlation, and the dashed line is the correlation for the randomized data (control).

In order to analyze and compare the topographical distributions of mHb and lHb clusters further, I generated correlation versus distance plots for each domain separately (Figure 3.24). In the plot for medial clusters, the pairwise correlation value drops to 0 around 500 µm, while for lateral clusters, the pairwise correlation value drops to 0 around 1000 µm. Furthermore, the peak pairwise correlation value is 0.3 for medial habenula clusters and 0.4 for lateral habenula clusters. These findings suggest that lHb neurons exhibit less well-defined spatial organization and more similar gene expression profiles compared to mHb neurons with more diverse cell types and stricter topography.

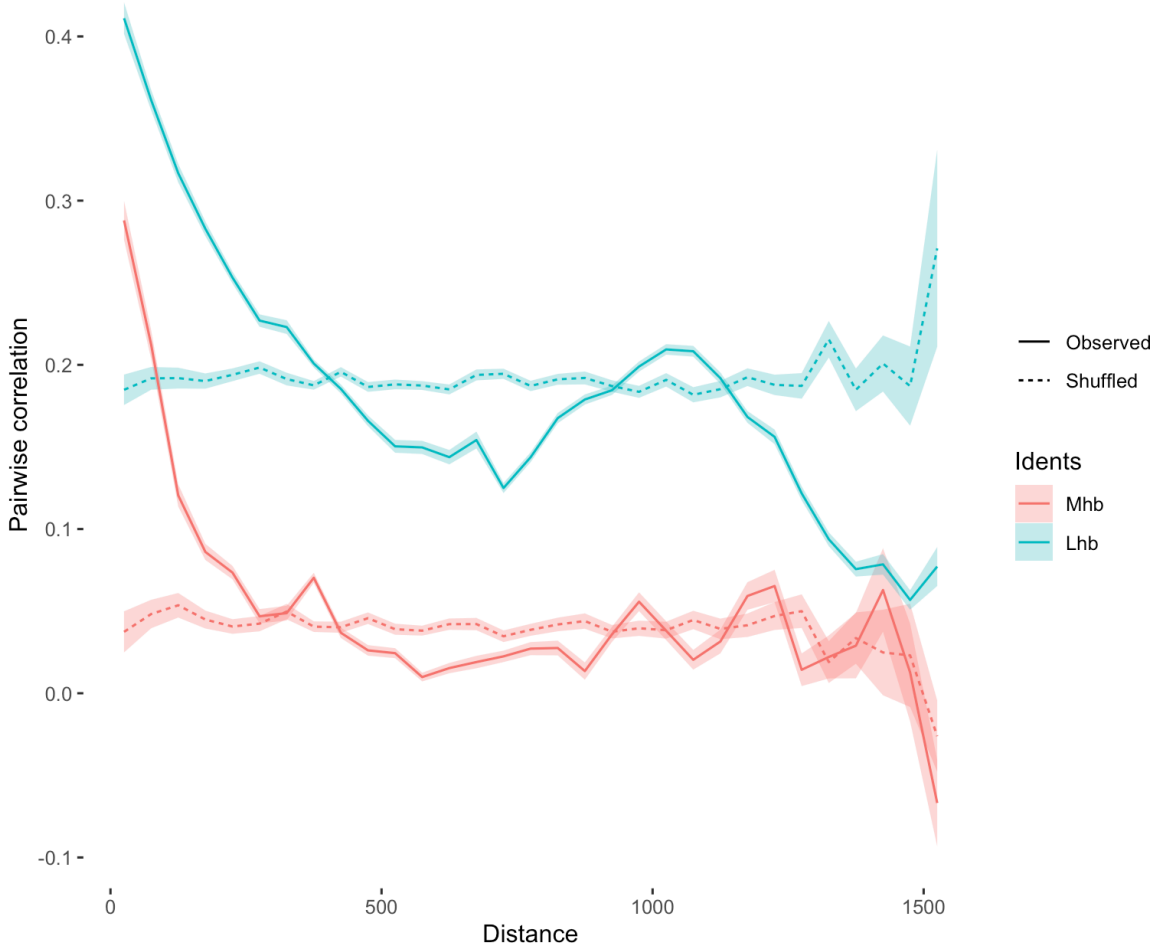

Figure 3.24: Correlation versus distance plots of the medial (red) and lateral (blue) habenula in the MERFISH dataset. The y-axis is the pairwise correlation value; the x-axis is the distance in µm. The solid line is the observed correlation, and the dashed line is the correlation for the randomized data (control). Distance is in µm.

The next objective is to cross-validate the spatial mouse data MERFISH by integrating and comparing it to the scRNA-seq mice data. Since scRNA-seq mice data has a

higher gene number than MERFISH data, this integration can contribute to the accuracy and reliability of the spatial data. Hence, combining these datasets can allow a more comprehensive understanding of gene expression patterns in their spatial context.

In order to integrate scRNA-seq mice data and spatial mouse data, I utilized the CoCy R package (Bredesen-Aa, 2024). As previously mentioned, CoCy utilizes Seurat's anchor-based Canonical Correlation Analysis (CCA) to integrate datasets and assess cluster similarity using the K Mutual Ratio (KMR) within their combined embedding (see methods). This tool generates a comparison matrix to analyze gene expression similarities and cluster overlaps between the MERFISH dataset and scRNA-seq data from adult mice (Figure 3.25).

The comparison heatmap between the adult mice single-cell RNA-seq data and the spatial dataset for mouse reveals a strong mutual correlation between lHb clusters and mHb clusters in both datasets (Figure 3.25). This indicates that the MERFISH data accurately reflects the cell types identified in the single-cell RNA-seq data, which validates the spatial dataset. Unexpectedly, there is a relatively low overlap between lHb cluster 2 from the single-cell RNA-seq and an mHb cluster 3 from the MERFISH. In addition, in the previous analysis, clusters 4 and 3 from the single-cell RNA-seq data were not expressing *tac2* or *pcdh10*, and their spatial location was unknown. In this comparison matrix, these clusters are finding matches with mHb clusters in the MERFISH data, suggesting these clusters might located in the mHb domain.

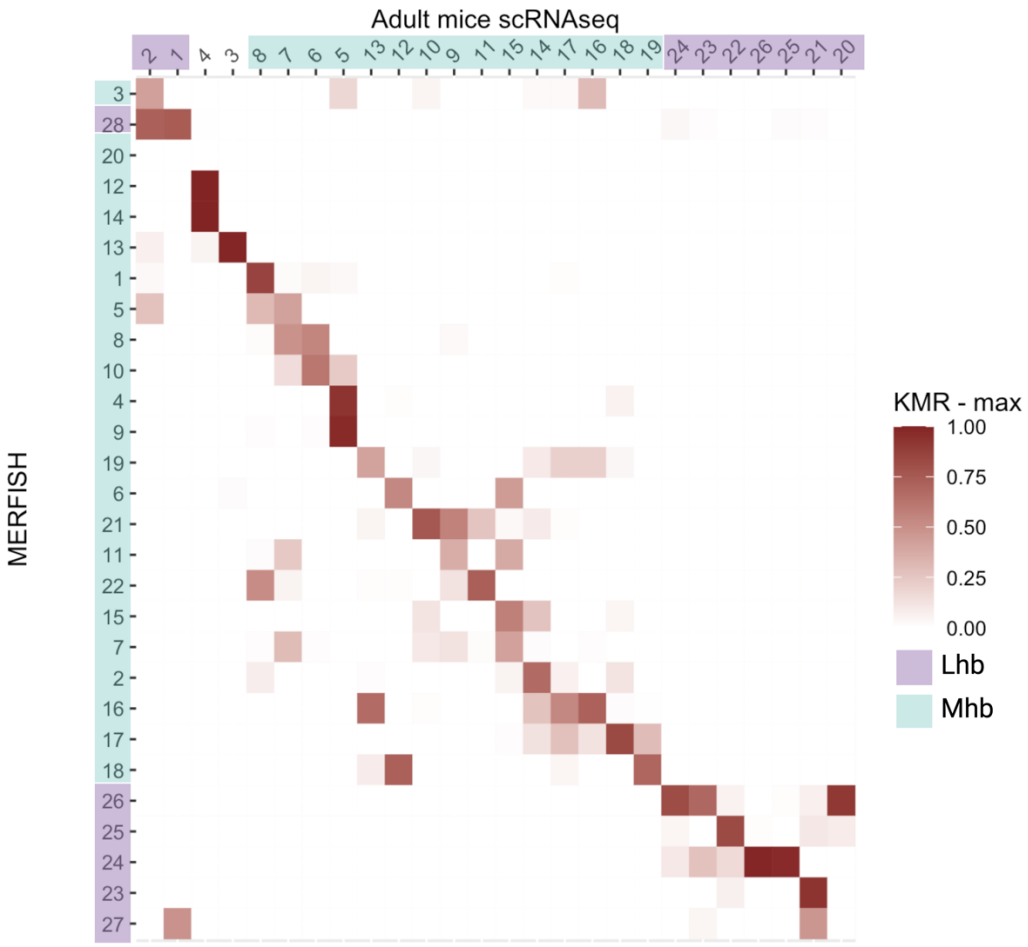

Figure 3.25: Comparison matrix of the MERFISH and mice scRNA-seq data. Columns are clusters of adult mouse scRNA-seq data, and rows are clusters of MERFISH data. The intensity of color shows a higher KMR value. Purple indicates lateral habenula cluster, and blue indicates medial habenula cluster. KMR: K-Mutual ratio.

3.5 Topographical Comparison of Adult Zebrafish Habenula and Adult Mouse Habenula

After identifying the topography with correlation versus distance plots of mouse and zebrafish, the next objective is to compare these species to understand how spatial gene expression patterns are organized. Furthermore, it will provide insights into how spatial organization differs or is conserved across species, indicating spatial organization (Figure 3.26). As previously discussed, a higher correlation is seen at closer distances in both zebrafish and mouse habenula graphs. In the mouse habenula, the decrease of correlation is

more gradual compared to the zebrafish habenula. This implies that the spatial organization is more dispersed in mouse habenula than in zebrafish habenula.

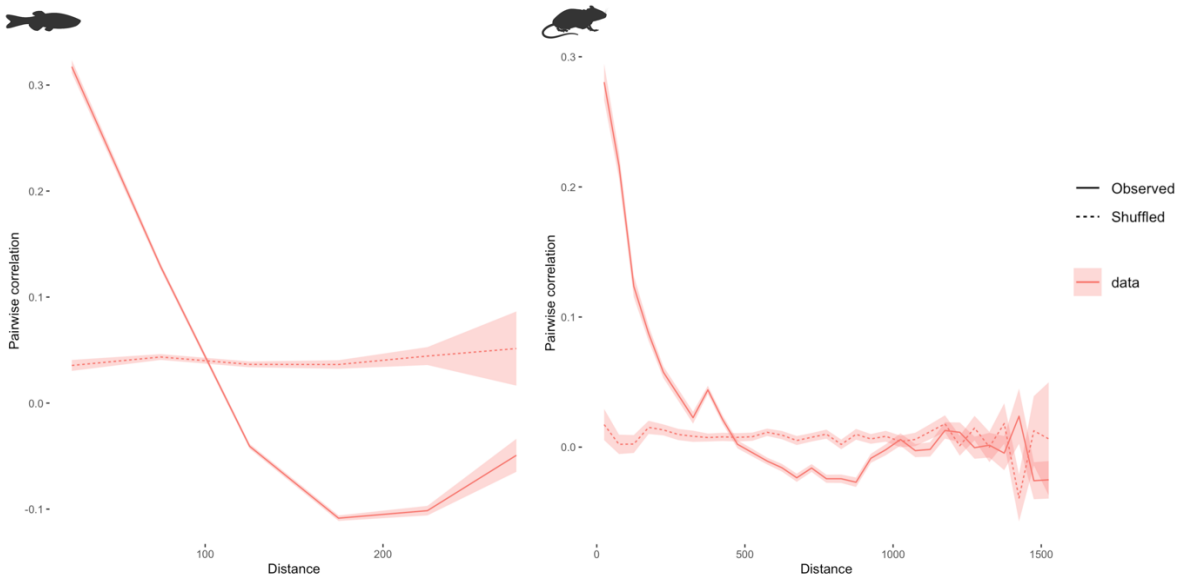

Figure 3.26: Correlation versus distance plot of (left) zebrafish habenula in the AZTEC (right) and mouse habenula in the MERFISH dataset. These plots show the pairwise correlation between gene expression levels with respect to distance within the habenula. The y-axis is the pairwise correlation value; the x-axis is the distance in um. The solid line is the observed correlation, and the dashed line is the correlation for the randomized data (control). (Created with BioRender).

3.6 Exploring Molecular Lateralization and A/symmetry of Adult Zebrafish and Adult Mouse Habenula

The asymmetry of habenula across species differs in terms of neurochemical composition, neuronal organization, and the extent of the asymmetry (Concha & Wilson, 2001; Ulrike Hüsken, 2014). Given the critical role that asymmetry plays in the functional organization of the habenula, such as sensory processing (Elena Dreosti, 2014), the next objective of the research is to study deeper into asymmetry and lateralization in adult zebrafish and mouse habenula. In order to analyze asymmetry, first, I separated the left and right domains of the habenula in both datasets, MERFISH and AZTEC (Figure 3.27).

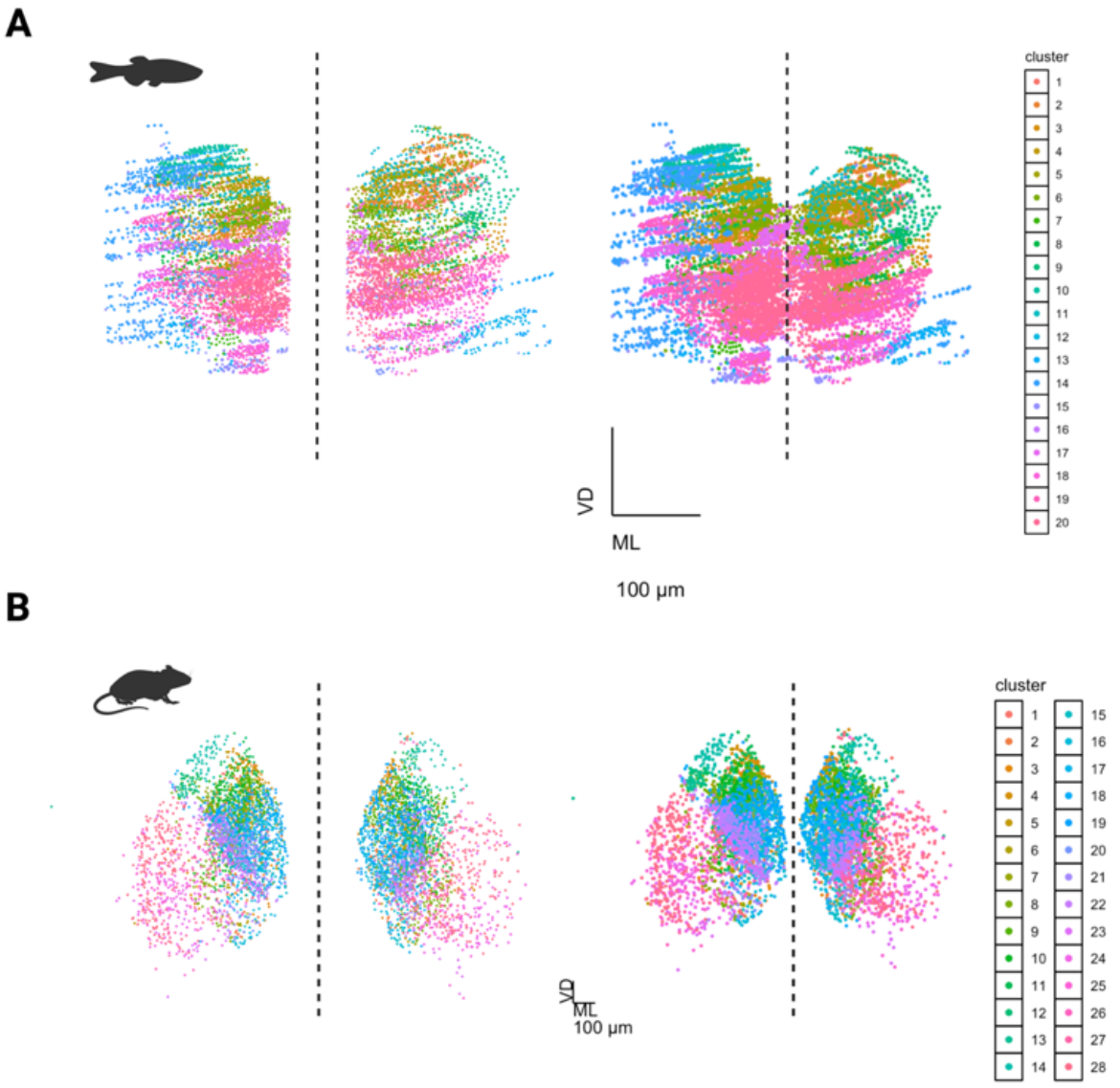

Figure 3.27: L-R separation of cell clusters in the habenula of zebrafish and mouse. (A) Coronal sections of (right) the L-R separated habenula of zebrafish, (left) whole zebrafish habenula. (B) The coronal section of the (right) L-R separated the habenula of the mouse and (left) the whole mouse habenula.VD: Ventral, dorsal. ML: Medial, lateral. Scale bars indicate 100 µm. (Created with BioRender).

After separating the two hemispheres, I analyzed the correlation vs distance plots for the left and right habenula of zebrafish and mouse, separately to understand hemispheric lateralization (Figure 3.28). In the right habenula of zebrafish, the correlation decreases more gradually and remains relatively higher over the same distances as the left habenula. This suggests that the right habenula in zebrafish has a more gradual spatial compared to the left habenula. On the other hand, the decrease of correlations in the left and right habenula of

mouse are relatively similar, indicating that the spatial organization in mouse habenula is more symmetric compared to zebrafish.

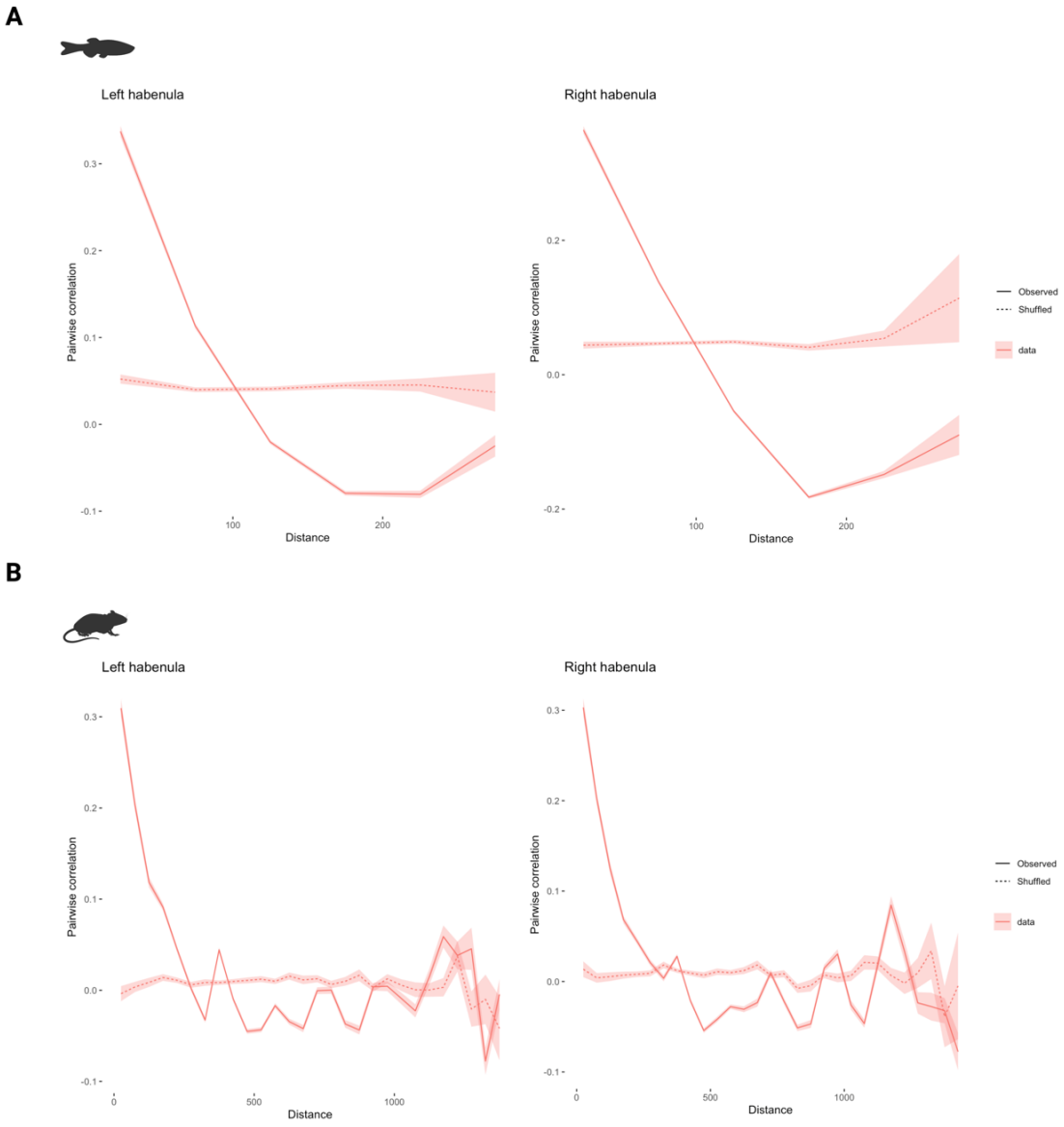

Figure 3.28: Correlation versus distance plot of L-R domains of the zebrafish habenula and mouse habenula. (A) Left and right domains of zebrafish habenula of AZTEC dataset. (B) Left and right domains of zebrafish habenula of MERFISH dataset. These plots show the pairwise correlation between gene expression levels with respect to distance within the habenula. The y-axis is the pairwise correlation value; the x-axis is the distance in um. The solid line is the observed correlation, and the dashed line is the correlation for the randomized data (control). (Created with BioRender).

Next, I calculated the distribution for the ratio of cells in each cluster of neurons in the left and right hemispheres, as shown in Figure 3.29. In the zebrafish habenula, there is a noticeable asymmetry in several clusters, such as clusters 12, 10, 14, and 17, which are predominantly left-dominated, and clusters 9, 1, and 2, which are predominantly right-dominated. Some clusters, such as 18, 5, and 11, show high symmetry. The distribution of cells between the right and left domains seems more symmetric in mouse habenula when compared to the zebrafish habenula. In mouse habenula, clusters like 26, 21, and 23 show a slight preference for the left hemisphere while clusters such as 27 and 16 show a higher percentage in the right hemisphere.

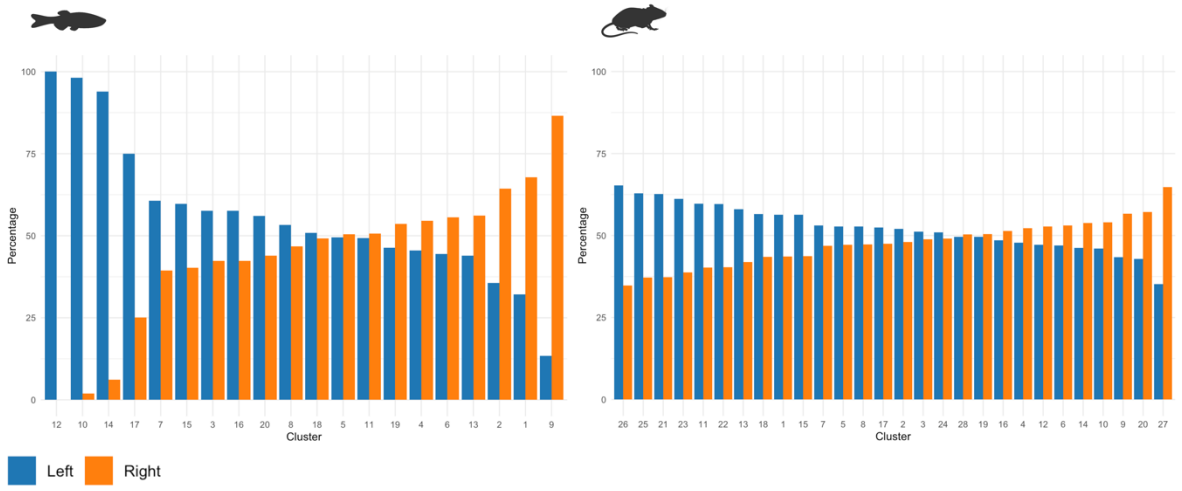

Figure 3.29: Asymmetry of clusters in the zebrafish and mouse habenula. Bar graph showing the percentage distribution of cells in the left (blue) and right (orange) domains of (A) zebrafish habenula and (B) mouse habenula. Each bar represents a distinct cell cluster, with the height indicating the proportion of cells on the left and right sides of the habenula. The x-axis denotes the cluster number, while the y-axis represents the percentage of cells. (Created with BioRender).

The observed asymmetry in clusters, calculated by the distribution percentages of cells on the left and right sides, provides an initial indication of lateralization. However, these percentages alone do not allow direct quantification and comparison of the Hb lateralization between mouse and zebrafish Hb. To overcome this limitation, the lateralization index (LI) for each cluster is calculated by the absolute value of the difference between the number of right cells and left cells, divided by the total number of cells (the sum of right and left cells). This calculation produces a value between zero (perfect symmetry) and one (perfect asymmetry). The lateralization index offers a more interpretable representation of

lateralization than directly comparing left and right percentages (see methods). In addition, the coordinates of each cell are shuffled with CoCy to create randomly distributed versions of AZTEC and MERFISH. These shuffled versions are used as controls (Figure 3.30). Furthermore, a significant lateralization threshold is calculated for zebrafish and mice separately (see methods).

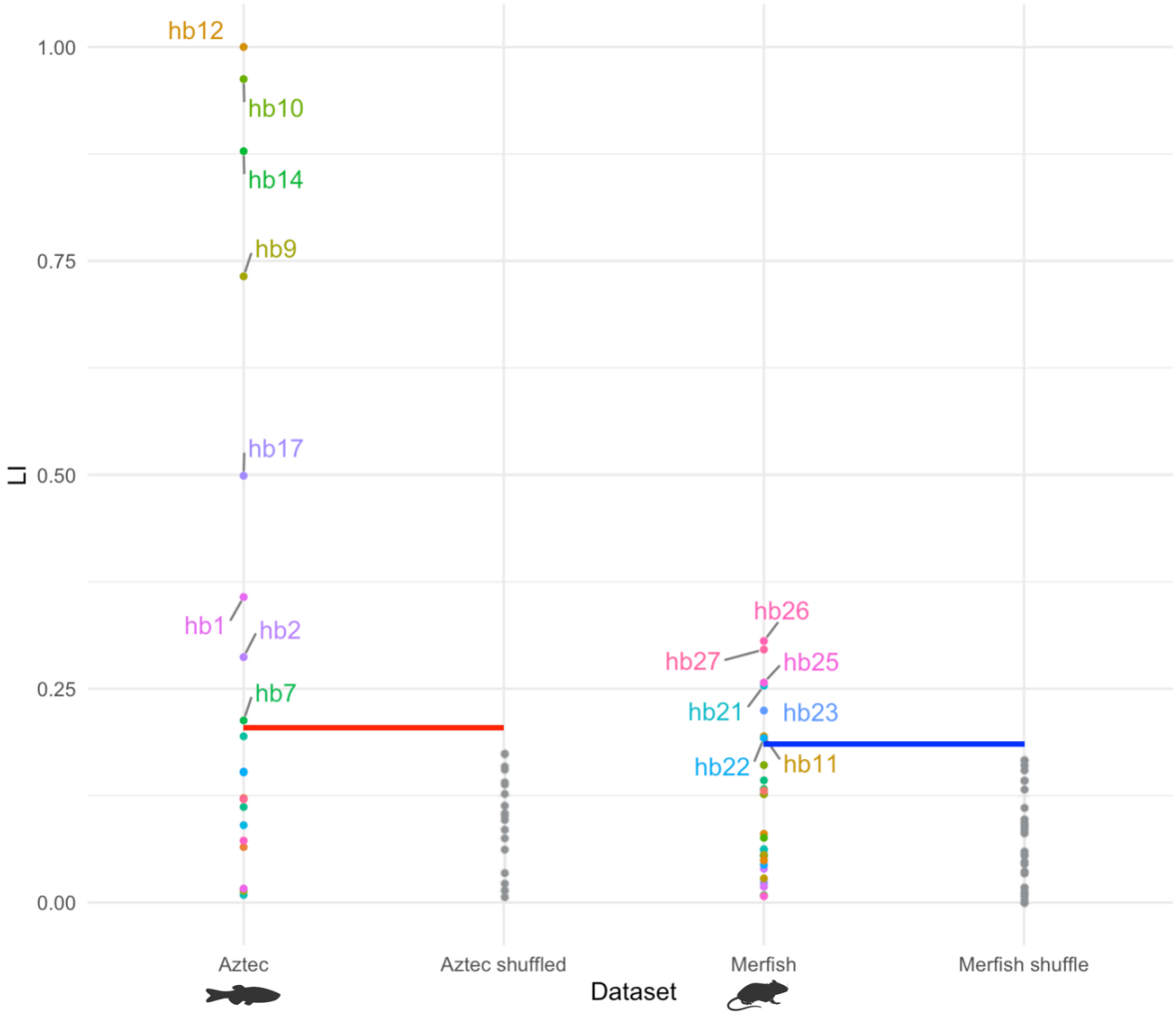

Figure 3.30: Lateralization of clusters in the zebrafish and mouse habenula. The y-axis represents the lateralization index (LI), and the x-axis shows the datasets used for the analysis, including the original and shuffled (control) versions for both AZTEC (zebrafish) and MERFISH (mouse) datasets. Each point represents a specific cell cluster, color-coded according to the cluster. (Created with BioRender).

Based on the thresholds, eight clusters in the AZTEC dataset show significant lateralization. These clusters are Dhb12, Dhb10, Ddhb14, Dhb9, Vdhb17, Dhb1, Dhb2, and

Vhb7. Similarly, seven clusters in the MERFISH dataset demonstrate significant lateralization. These clusters are Ldhb26, Ldhb27, Ldhb25, Mhb21, Ldhb23, Mhb22, and Mhb11(Figure 3.26). Notably, the lateralization index is significantly lower in mouse Hb clusters compared to zebrafish Hb clusters. The significant lateralization index for zebrafish Hb clusters ranges between 1 and 0.21, while for mouse Hb clusters, it ranges between 0.30 and 0.19.

Lastly, I aim to calculate the lateralization of individual genes in both the zebrafish and mice datasets in order to analyze the gene expression asymmetry in the habenula across different species. The lateralization index is calculated using the same formula for cluster lateralization. However, instead of right and left cells, count numbers of right and left cells are used (see methods). Figure 3.31 shows that the gene lateralization is significantly higher in the zebrafish habenula than in mouse habenula. Even the most lateralized genes in the mouse habenula exhibit a relatively symmetrical expression profile.

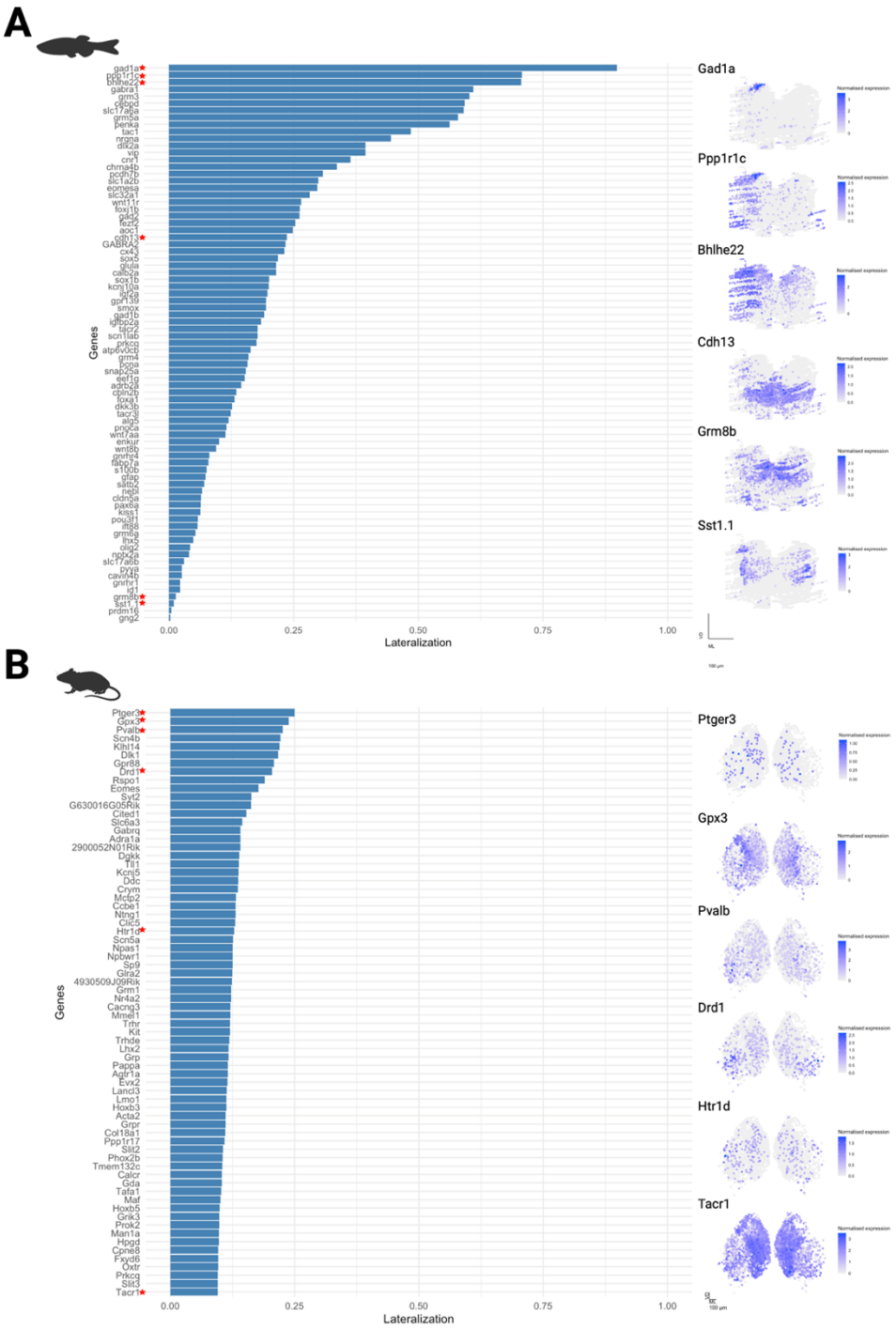

Figure 3.31: The plot of gene lateralization in the zebrafish and mouse habenula. The y-axis represents the gene names, and the x-axis shows the lateralization index value. On the left side, the gene lateralization graph of (A) zebrafish and (B) mouse habenula is shown. Red stars indicate the genes selected as examples to show. Coronal sections of the spatial distribution of example genes are shown on the right side. The bar indicates the normalized expression (from blue to gray, highest to lowest, respectively.) The scale bar is 100 μm. VD: Ventral, dorsal. LM: Lateral, medial. (Created with BioRender).

3.7 Conservation of Cell Groups Among Adult Zebrafish Habenula and Adult Mouse Habenula

The next objective is to study the conservation of cell groups. scRNA-seq datasets and spatial datasets are compared between zebrafish and mouse habenula. For this reason, CoCy and KMR-based comparison matrices are constructed (see methods) (Bredesen-Aa, 2024). 7701 common genes are found between scRNA-seq zebrafish and mouse datasets with EggNOG mapping.

Zebrafish scRNA-seq clusters 1 and 2 exhibit a strong correlation with nearly all lHb clusters identified in the mouse scRNA-seq dataset (Figure 3.32). The clusters 1 and 2 of zebrafish did not show *aoc1* expression, but the integration with the AZTEC dataset revealed that these clusters might be ventral. Interestingly, almost all mouse lHb clusters correlate with the only non-*aoc1* expressing ventral clusters in zebrafish habenula. Dorsal clusters in zebrafish habenula only show six strong overlaps with mouse mHb clusters. Both dHb clusters 6 and 3 overlap with mHb cluster 7. Dorsal cluster 13 shows a correlation with mHb cluster 6. Notably, cluster 4 in the mouse dataset, which is predicted to be an mHb cluster based on the comparison with the MERFISH dataset, demonstrates a significant correlation with three distinct dHb (12, 11, and 5) clusters in the zebrafish scRNA-seq dataset (Figure 3.32). Unexpectedly, most of the *aoc1* expressing vHb clusters in the zebrafish dataset show correlation with mHb clusters in the mouse dataset, with the exception of vHb clusters 16 and 19.

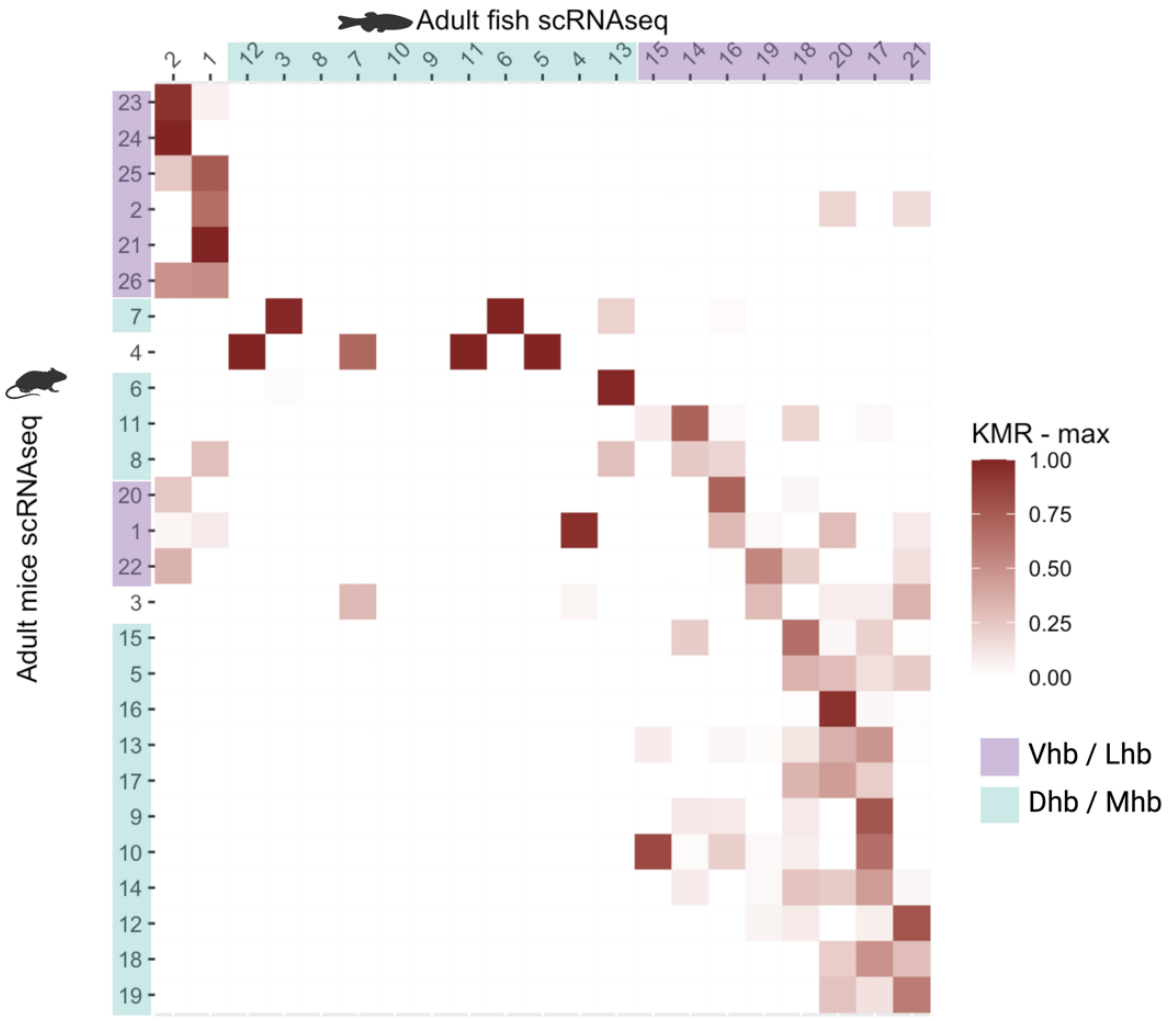

Figure 3.32: Comparison matrix of the zebrafish and mouse scRNA-seq data. Columns are clusters of the zebrafish scRNA-seq data; rows are clusters of the mice scRNA-seq data. The intensity of color shows a higher KMR value. Purple color indicates lateral and ventral habenula clusters, and blue indicates medial and dorsal habenula clusters. KMR: K-Mutual ratio.

The spatial datasets AZTEC and MERFISH share a limited number of common genes, with only 27 overlapping genes. This limited overlap accounts for the lower level of conservation observed compared to scRNA-seq datasets. Despite this limitation, a general correlation is visible between zebrafish and mouse.

There is a significant overlap between dHb cluster 12 in the zebrafish habenula and mHb cluster 8 in the mouse habenula (Figure 3.33). Additionally, an interesting correlation is between vHb cluster 16 in zebrafish and mHb cluster 2 in mouse. There is a relatively high overlap between dHb cluster 5 and mHb cluster 13 and between dHb cluster 9 and mHb

cluster 6. Moreover, it is seen that most dHb clusters in the zebrafish habenula are correlated with mHb clusters in the mouse habenula. Surprisingly, the vHb clusters in the AZTEC dataset also show correlations with mHb clusters. The only exception to this result is two ventral clusters (7 and 13) that do not show expression for *aoc1*. These clusters are the only ventral cell groups that show their highest overlap with the lHb clusters in mouse.

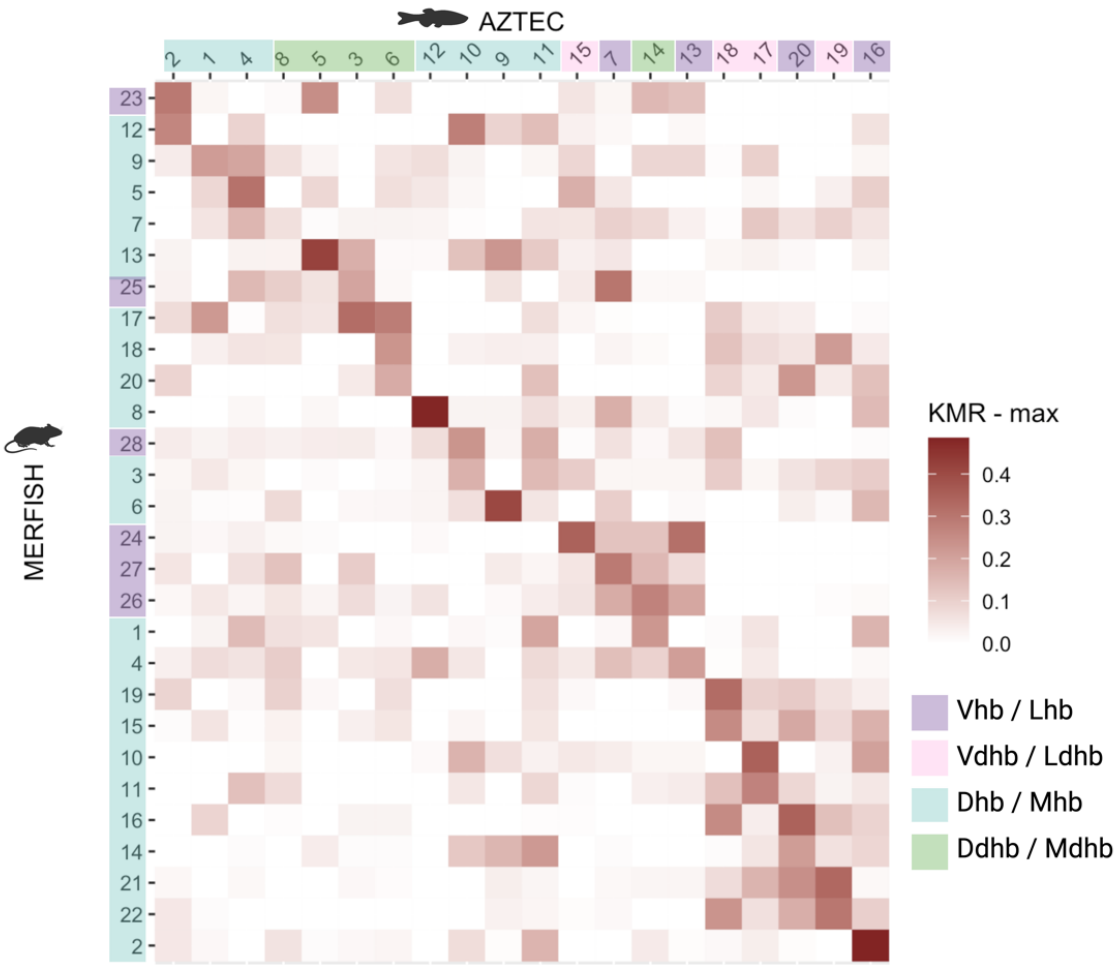

Figure 3.33: Comparison matrix of the AZTEC and MERFISH datasets. Columns are clusters of AZTEC data; rows are clusters of MERFISH data. The intensity of color shows a higher KMR value. Purple color indicates ventral habenula cluster, pink indicates ventral dominated habenula cluster (meaning ventral cluster with some cells in the dorsal domain), blue indicates dorsal habenula cluster; green indicates dorsal dominated habenula cluster (meaning dorsal cluster with some cells in the ventral domain). KMR: K-Mutual ratio.

Chapter 4:
DISCUSSION

4.1 Diversity and Topografy Cell Types in Vertebrate Habenula

Few recent studies have been published about the cell heterogeneity of habenula (Hashikawa et al., 2020; Pandey et al., 2018). However, a detailed analysis of each cell group and their functional and topographical organization remained to be addressed. Pandey et al. identified 17 different cell types in the adult zebrafish habenula (Pandey et al., 2018). Additionally, Hashikawa et al. show that the mouse habenula consists of 12 different clusters, 6 mHb and 6 lHb, while Wallace et al. described 9 cell types in mouse habenula, 5 mHb and 4 lHb (Hashikawa et al., 2020; Wallace, 2020). All these studies are relatively recent, and were conducted with scRNA-seq datasets combined with in situ hybridization experiments. Given the crucial role of habenula, I believe a more detailed analysis of distinct cell types is needed. Hence, my thesis was an attempt to dissect spatio-molecular features of vertebrate habenula in more detail in two different species.

Despite histological staining in previous studies (Calvigioni D, 2023; Hashikawa et al., 2020; Lazaridis et al., 2019; Pandey et al., 2018), spatio-molecular organization of adult mouse and zebrafish habenula remains to be explored. Specifically, differences and similarities between the spatial organization of cells in dHb – vHb and mHb – lHb remain unknown. When we consider the distinct functions of these regions, a spatial organization comparison would be beneficial to understand their respective roles better and elucidate the underlying mechanisms governing their function.

By using an iterative tool, clustree, for visualizing the correct resolution for the right number of molecular clusters, I identified 21 and 20 different clusters in single cell RNA-seq and spatial-transcriptomic datasets of zebrafish habenula. For adult mice habenula, I identified 26 distinct cell groups in scRNA-seq and 28 distinct cell groups in the MERFISH spatial-transcriptomic dataset. The fact that cell group numbers are relatively similar within the two datasets of the same species lends confidence to the accuracy of the analysis and the selected resolution parameter when identifying clusters.

In the adult zebrafish scRNA-seq dataset, based on ventral markers *aoc1* and *kiss1*, eight cell types (clusters 14 - 21) are ventral, and eleven cell types (clusters 3-13) are dorsal. Two clusters, 1 and 2, did not show any expression of typical ventral or dorsal markers. These clusters' anatomical location is predicted as ventral in comparison with the spatial-transcriptomic dataset, AZTEC (Figure 3.16). Cluster 2 showed strong and specific expression of *gad2*, *gad1b*, and *slc32a1*, suggesting that this might be the only gabaergic cell group in this dataset. In the UMAP visualization, dorsal clusters and ventral clusters seem to create chunks within themselves, whereas clusters 1 and 2 seem separate from both of the groups (Figure 3.4), suggesting that these clusters are molecularly distant to both of the groups, which I also confirmed by inspecting the heatmap of gene expression patterns of these clusters (Figure 3.4). The molecular similarity dendrogram indicates that ventral clusters are molecularly similar to each other, and the same is true within dorsal clusters. In ventral clusters, it is seen that each gene is expressed in multiple ventral clusters, and the expressed genes are differentiated among ventral clusters not based on expressed or not expressed but based on the relative levels of expression. There are some genes, however, showing a more specific expression in some ventral clusters, such as *h1fx* (cluster 21) and *egr4* (clusters 15 and 18).

On the other hand, the genes expressed in dorsal clusters showed a more distinct expression pattern among clusters, with high specificity to individual clusters. In the heatmap of gene expression patterns of these clusters, it is seen that the gene expression profile in dorsal clusters is more unique and heterogeneous when compared to the ventral clusters' gene expressions (Figure 3.4). This suggests that there is a greater diversity of gene expression patterns and distinct subcircuits within dorsal Hb when compared to ventral Hb.

In the scRNA-seq mice dataset, clusters 1, 2, and 20-26 are predicted to be located in lHb due to the expression of *pcdh10*, and clusters 5-19 are predicted to be located in mHb due to the expression of *tac2*. Clusters 3 and 4 did not show any expression of either of the genes and were predicted to be mHb clusters in comparison with the spatial dataset MERFISH (Figure 3.25). In the UMAP visualization, medial and lateral clusters form distinct groups, while clusters 3 and 4 appear to be separate from both (Figure 3.8). This separation suggests that clusters 3 and 4 are molecularly distinct from the dorsal and ventral groups.

This observation is further supported by the heatmap analysis of gene expression patterns, confirming these clusters' molecular divergence (Figure 3.8). Similar to the zebrafish habenula, the heatmap of the scRNA-seq mice dataset also showed a more shared gene expression profile in lHb clusters and a more distinct distribution in mHb clusters.

This pattern is also seen in the AZTEC dataset, where *aoc1* expressing five cell groups (clusters 16-20) show a more scattered distribution, and the remaining clusters (1-15) show more distinct expression profiles. Interestingly, the anatomical distribution of clusters reveals that clusters 7 and 13 are also located in the ventral habenula, which did not have *aoc1* and *kiss1* expression. These two clusters also show relatively unique gene expression distribution, indicating that the homogeneity in gene expression might not be seen in all ventral clusters but only in *aoc1*-expressing ventral clusters. Furthermore, this pattern has also been seen in the MERFISH dataset: lHb cell groups (clusters 23-28) show a more homogenous gene expression pattern compared to the mHb cell groups. This suggests that dHb/mHb clusters have a more distinct set of genes being expressed at different levels, indicating specialized functions or roles for these clusters.

On the contrary, gene expression distribution in vHb/lHb clusters shows less variability and more uniformity. This indicates that ventral clusters might have a more shared function. Clusters that are not *aoc1* expressing seem to be an exception to this pattern, suggesting that non-*aoc1* expressing ventral clusters might not have a commonly shared function as other ventral clusters.

The spatial distribution of cell groups revealed that there is a strong spatial grouping between the cells that are molecularly similar. Interestingly, cluster 15, which is a GABAergic cluster, did not fit into this interpretation since it shows a more scattered cell distribution in the habenula. The distribution of cluster 15 might reflect a unique role for this GABAergic cluster in modulating activity throughout the habenula.

Interestingly, the *tac2* mHb marker gene did not show expression in clusters 12, 13, and 14, which are shown to be located in the mHb, revealed by spatial distribution analysis. This analysis also showed a high spatial organization in the mouse habenula. Correlation vs distance plot also showed that spatially closer cells are more molecularly similar, which

further verified spatial organization. In order to further verify the heatmap results for lHb and mHb gene distribution profile differences, correlation vs distance plots of lHb and mHb cells are constructed. These plots revealed that lHb clusters indeed exhibit more similar gene expression profiles compared to mHb clusters.

The correlation of gene expression as a function of distance plot (Figure 3.14) further showed and quantified the spatio-molecular organization of the zebrafish habenula. This plot showed that as the distance increases between pairs of cells, their molecular identities also become more different. Additionally, the correlation vs distance plots for dorsal and ventral habenula further supported the results from the heatmap, where the initial correlation value in ventral clusters was higher than the dorsal clusters. Furthermore, as the distance increases, ventral cells exhibit a slower decline in correlation, suggesting higher gene expression similarity among spatially distant ventral cells.

Overall, this analysis showed that the habenula consists of many molecularly distinct cell types, indicating a high degree of cellular diversity and specialization within this brain region in both zebrafish and mouse. In addition, high spatial organization is observed in both zebrafish and mouse habenula, indicating a conserved role of spatial arrangement in the habenula across species. Also, it is seen that *aoc1* expressing ventral clusters have a more homogeneous gene expression when compared to dorsal clusters in the zebrafish habenula. Furthermore, the same pattern is observed in the homologous lHb and mHb clusters in the mouse habenula, indicating a conserved organizational principle across these species.

4.2 Asymmetry of Zebrafish and Mouse Habenula

As it is discussed in the introduction section, most vertebrates have some degree of asymmetry in the habenula region. However, degree of these asymmetries varies widely (Concha & Wilson, 2001; Ulrike Hüsken, 2014). The asymmetry of the zebrafish habenula has been extensively studied, whereas there is no strong evidence about the asymmetries of the mouse habenula. I focused on the asymmetry and lateralization of the habenula in both species to identify the asymmetrical or symmetrical profiles of habenular domains (such as ventral or lateral), cell groups, and gene expression.

Furthermore, the left and right cell distributions of each cluster of the zebrafish habenula revealed highly asymmetric clusters. However, interestingly, there are highly symmetrical clusters as well. In the mouse habenula, this distribution seems more symmetrical, although there are some clusters showing low-degree asymmetry. In the zebrafish habenula spatial images, a clear asymmetry is seen in several cell groups, such as clusters dHb9, dHb10, dHb12, and dHb14. On the other hand, in the mouse habenula, cell groups seem to distribute symmetrically between left and right. Thanks to the lateralization index and the threshold set for significant lateralization, the analysis went further than just predicting which clusters are highly asymmetric.

The lateralization plot reveals that there are 8 highly asymmetric clusters in the zebrafish habenula and 7 mildly but significantly asymmetric clusters in the mouse habenula. In the AZTEC dataset, these clusters are categorized as 6 dorsal (dHb12, dHb10, dHb14, dHb9, dHb1, dHb2) and 2 ventral (vHb17, vHb7). In the MERFISH dataset, the asymmetric clusters are categorized as 4 lateral (lHb26, lHb27, lHb25, lHb23) and 3 medial (mHb21, mHb22, mHb11). It is important to note that even though 7 clusters are found to be significantly lateralized, their lateralization degrees are significantly lower than zebrafish, suggesting that zebrafish habenula is more lateralized than mouse habenula. The gene expression patterns also parallel this result. The most lateralized genes in the mouse habenula have a lateralization index value of around 0.25, whereas in the zebrafish habenula there are several highly lateralized genes with lateralization values of more than 0.5. This highlights that zebrafish habenula exhibits more asymmetry in gene expression compared to the mouse habenula.

4.3 Conservation of Habenula Accros Species

There is plenty of evidence that the dorsal and ventral habenula of zebrafish are homologous of the mammalian medial and lateral habenula, respectively (Amo et al., 2014; Bianco & Wilson, 2009). Furthermore, a relatively recent study showed that the habenula is transcriptionally conserved among zebrafish and mouse by comparing the gene expressions of scRNA-seq datasets of these two species (Hashikawa et al., 2020). The researcher also created an integrated dataset and compared it with zebrafish and mouse, separately. In this

project, I followed a different analytical methodology to compare these two species directly with each other. Furthermore, I also compared the spatial datasets to see the conservation of spatial organization and topography. Lastly, I studied whether there is a conservation between lateralized clusters of two species. The results of this study show that there is a homology between mHb and dHb, as previous studies suggested. However, the comparison matrices between zebrafish and mouse habenula reveal a conservation between mHb and vHb, which is not previously discussed.

7701 common genes are found between scRNA-seq zebrafish and mice dataset, increasing the comparison accuracy. Zebrafish scRNA-seq clusters 1 and 2, which overlap with the ventral AZTEC clusters, exhibit a strong correlation with nearly all lHb clusters identified in the mice scRNA-seq dataset. Notably, cluster 4, which overlap with the mHb clusters in MERFISH, in the mouse scRNA-seq dataset demonstrates a significant correlation with three distinct dHb clusters in the zebrafish scRNA-seq dataset. In addition, only a part of the dHb clusters in the zebrafish scRNA-seq dataset are correlated with some mHb clusters in the mouse scRNA-seq dataset. Unexpectedly, we observe that most of the vHb clusters in the zebrafish scRNA-seq dataset also show correlation with mHb clusters in the mouse scRNA-seq dataset.

The spatial mouse and zebrafish datasets share a limited number of common genes, with only 27 genes overlapping between them. The small number of common genes between the spatial datasets (AZTEC and MERFISH) can limit the robustness of comparative analyses. Hence, there might be more correlated clusters that cannot be identified with this comparison matrix. In this comparison matrix, four highly correlated overlaps are found, three of them are between dHb and mHb clusters. Surprisingly, the highest KMR value overlap is between vHb cluster 16 and mHb cluster 2. There is no significant correlation between lateralized clusters of zebrafish habenula and lateralized clusters of mouse habenula, suggesting that these lateralized clusters are not evolutionarily conserved.

The results of this study demonstrate a homology between the medial habenula (mHb) and the dorsal habenula (dHb), consistent with previous findings in the literature. This suggests that these regions share common features in terms of structure and function. However, our comparative analyses between the zebrafish and mouse habenula reveal a novel

conservation between the medial habenula (mHb) and the ventral habenula (vHb), a relationship that has not been previously discussed or highlighted in the scientific community. This unexpected finding suggests that there may be previously unrecognized evolutionary and functional connections between these regions across species. This variability might be crucial for the specialized roles that different parts of the habenula play in modulating behavior and neural processes.

4.4 Verification of Spatial Transcriptomic by Integration to Single-cell, and HCR

In this project, by using one scRNA-seq and one spatial dataset for each species, we aim to benefit from different advantages and overcome the limitations these technologies have. ScRNA-seq datasets have high gene numbers; however, they lack spatial context. They are useful in discovering distinct cell types and their gene expression profiles; however, they remain to reveal the spatial organization and interactions between cells within the tissue. Spatial datasets, on the other hand, provide significant spatial context and allow us to understand the physical locations and interactions between cells, but they typically have a lower number of genes. Consequently, they offer valuable insights into the spatial organization of gene expression. However, they may miss some gene expression details due to the limited number of genes. By utilizing both technologies, I aim to have a more comprehensive understanding of the cell types and topography of the habenula. Furthermore, scRNA-seq and spatial datasets from the same species are integrated for cross-verification and validation in order to ensure that observed cluster patterns are accurate and not artifacts of a single experimental approach. I identified the potential spatial location of two clusters of scRNA-seq mice that don't have any expression for marker genes. These clusters, 3 and 4, both overlapped with the different mHb clusters in MERFISH, suggesting that these clusters might be located in mHb. Similarly, the potential anatomical distribution of two clusters (1 and 2) of scRNA-seq zebrafish is predicted by integrating it with AZTEC data. This integration revealed that these clusters might be located in the ventral domain. Overall, this approach benefits from the strengths of both technologies, which results in more accurate interpretations of the datasets.

To further validate the anatomical distribution of gene expression patterns in the AZTEC dataset, I performed the hybridization chain reaction (HCR) with two genes of particular interest to our laboratory, *grm6a*, and *aoc1*, which are predominantly expressed in the dorsal and ventral domains of the zebrafish habenula, respectively. The gene expression profiles observed for *aoc1* and *grm6a* closely matched the patterns obtained from the AZTEC spatial dataset. This alignment between the HCR results and the AZTEC dataset provides further evidence for the reliability and accuracy of the AZTEC dataset in representing the spatial distribution of gene expression within the zebrafish habenula.

4.5 Relationship of Molecular and Functional Topography in Vertebrate Habenula

As previously mentioned, a study revealed most neurons from the left habenula connect to the dorsal part of the interpeduncular nucleus (IPN) (Ulrike Hüsken, 2014). Left-enriched dorsal clusters 10 and 12 might belong to this group. In addition, neurons from the right habenula connect to the ventral part of the IPN in the zebrafish. (Ulrike Hüsken, 2014). From this, right-enriched dorsal cluster 9 might belong to this group.

Also, in 4-days-old zebrafish, researchers have seen that neurons responding to light are located on the left side of the habenula, and neurons responding to odor are more commonly found on the right side (Elena Dreosti, 2014). Even thought the AZTEC dataset is adult zebrafish, the left and right distribution of clusters might be associated with these neurons.

Previous studies show that zebrafish habenula's spontaneous activity has a topographical organization (Bartoszek, 2021; Jetti et al., 2014). In zebrafish and mouse spatial datasets, high spatial organization is seen between the molecularly similar cell groups (Figure 3.26). Future functional imaging in adult zebrafish and mouse habenula could enable the mapping of functional topographies in alignment with spatio-molecular habenular organizations. Considering the distinct neuron cell types found in the zebrafish and mouse habenula, integrating these findings with functional imaging can allow for the precise mapping of functional topographies and their correlation with spatio-molecular organizations.

Chapter 5:
CONCLUSION

In conclusion, I identified several distinct cell groups in both zebrafish and mouse habenula. Also, I generated a detailed gene expression profile of each cell groups of both species. This can be useful for understanding the molecular mechanisms of different habenular cell types and their function. Then, I showed the differentiation within the dHb/mHb and vHb/lHb cell groups. Specifically, ventral groups showed more homogeneous gene expression profiles in the zebrafish habenula, whereas dorsal groups showed greater diversity and uniqueness in gene expression. Similarly, in the mouse habenula, lateral cell groups show more homogeneity, while medial cell groups show more diverse gene expression.

The findings of this study also revealed a high spatial organization in both species. Notably, the vHb/lHb clusters have broader spatial organization and more similar gene expression profiles compared to the dHb/mHb clusters in both species. This is important since it suggests that the vHb/lHb clusters may share a more common function in habenula. Understanding these differences of spatial organization and gene expression can provide deeper insights into the functional specialization within the habenula and its evolutionary conservation across species.

Additionally, the results show that the lateralization of zebrafish habenula is significantly higher than mouse habenula. However, some clusters in the mouse habenula also exhibit a degree of lateralization. This finding underscores the evolutionary differences in brain lateralization mechanisms between species.

Lastly, I identified potentially conserved cell groups between the zebrafish and mouse habenula. Medial clusters mostly overlapped with the dorsal clusters and non-*aoc1* expressing ventral clusters are found to be overlapping with lateral clusters. Unexpectedly, *aoc1* expressing ventral clusters overlapped with the medial clusters. These findings are significant as they support the conservation of the habenula between zebrafish and mouse, but also they indicate that the overall homology accepted between mouse lateral habenula and zebrafish ventral habenula might not be as definitive as suggested previously.

Overall, this project aims to enhance the understanding of the cellular diversity and spatial organization within the habenula by combining scRNA-seq data with spatial datasets. This project can contribute to understanding the habenula's cell type heterogeneity, and spatial organization in different species. Furthermore, this project provides insights about the evolutionary conservation defree of the habenula in two different species. Additionally, identifying the lateralization of cell groups of both zebrafish and mouse habenula can shed light on the functional asymmetry and specialization. By understanding cell type heterogenity, conservation, spatial organization and lateralization we can gain a deeper insight into how the habenula functions and its evolutionary significance.

Chapter 6:

REFERENCES

Chapter 7:

Adil, A., Kumar, V., Jan, A. T., & Asger, M. (2021). Single-Cell Transcriptomics: Current Methods and Challenges in Data Acquisition and Analysis. *Front Neurosci*, *15*, 591122. https://doi.org/10.3389/fnins.2021.591122

Amo, R., Aizawa, H., Takahoko, M., Kobayashi, M., Takahashi, R., Aoki, T., & Okamoto, H. (2010). Identification of the zebrafish ventral habenula as a homolog of the mammalian lateral habenula. *Journal of Neuroscience*, *30*, 1566-1574. https://doi.org/10.1523/JNEUROSCI.3690-09.2010

Amo, R., Fredes, F., Kinoshita, M., Aoki, R., Aizawa, H., Agetsuma, M., Aoki, T., Shiraki, T., Kakinuma, H., Matsuda, M., Yamazaki, M., Takahoko, M., Tsuboi, T., Higashijima, S., Miyasaka, N., Koide, T., Yabuki, Y., Yoshihara, Y., Fukai, T., & Okamoto, H. (2014). The habenulo-raphe serotonergic circuit encodes an aversive expectation value essential for adaptive active avoidance of danger. *Neuron*, *84*(5), 1034-1048. https://doi.org/10.1016/j.neuron.2014.10.035

Bartoszek, E. M., Ostenrath, A. M., Jetti, S. K., Serneels, B., Mutlu, A. K., Chau, K. T. P., & Yaksi, E. (2021). Ongoing habenular activity is driven by forebrain networks and modulated by olfactory stimuli. *Current Biology*, *31*, 3861-3874. https://doi.org/https://doi.org/10.1016/j.cub.2021.08.021

Benedict Nilges , I. C., Richa Das, Andreas Geipel , Christian Korfhage , Frank Reinecke , Madeline A. Lancaster , Nachiket Kashikar (Resolve Biosciences). Molecular Cartography™ of brain organoids to unravel the spatial logic of neural tissue development. *https://resolvebiosciences.com*.

Bianco, I. H., & 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*(1519), 1005-1020. https://doi.org/10.1098/rstb.2008.0213

Binaree, I. HRTC Protocol (Code of Protocol: C1001, Revision of Protocol: 181131130). *Retrieved from* *http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf*.

Bredesen-Aa, B. A., Hinrichsen, F., ..., Yaksi, E. . (2024). Single-cell volumetric transcriptomic atlas of the adult zebrafish forebrain reveals teleost homologues of cortical and subcortical structures from terrestrial vertebrates. *manuscript in preparation*, *unpublished*.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Seurat: Tools for Single Cell Genomics (Version 4.0.0) [Computer software]. . *Retrieved from* *https://cran.r-project.org/package=Seurat*.

Calvigioni D, F. J., Le Merre P, Slashcheva M, Jung F, Ortiz C, Lentini A, Csillag V, Graziano M, Nikolakopoulou I, Weglage M, Lazaridis I, Kim H, Lenzi I, Park H, Reinius B, Carlén M, Meletis K. (2023). Esr1+ hypothalamic-habenula neurons shape aversive states. *Nat Neurosci*, *26*, 1245-1255. . https://doi.org/10.1038/s41593-023-01367-8

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development*, *145*(12). https://doi.org/10.1242/dev.165753

Choudhary, S., & Satija, R. (2022). Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol*, *23*(1), 27. https://doi.org/10.1186/s13059-021-02584-9

Cohen, S. R., & Melzack, R. (1993). The habenula and pain: repeated electrical stimulation produces prolonged analgesia but lesions have no effect on formalin pain or morphine analgesia. *Behav Brain Res*, *54*(2), 171-178. https://doi.org/10.1016/0166-4328(93)90076-3

Concha, M. L., & Wilson, S. W. (2001). Asymmetry in the epithalamus of vertebrates. *J Anat*, *199*(Pt 1-2), 63-84. https://doi.org/10.1046/j.1469-7580.2001.19910063.x

Contestabile, A., & Flumerfelt, B. A. (1981). Afferent connections of the interpeduncular nucleus and the topographic organization of the habenulo-interpeduncular pathway: an HRP study in the rat. *J Comp Neurol*, *196*(2), 253-270. https://doi.org/10.1002/cne.901960206

D'Gama, P. P., & Jurisch-Yaksi, N. (2023). Methods to study motile ciliated cell types in the zebrafish brain. *Methods Cell Biol*, *176*, 103-123. https://doi.org/10.1016/bs.mcb.2023.01.020

deCarvalho, T. N., Subedi, A., Rock, J., Harfe, B. D., Thisse, C., Thisse, B., Halpern, M. E., & Hong, E. (2014). Neurotransmitter map of the asymmetric dorsal habenular nuclei of zebrafish. *Genesis*, *52*(6), 636-655. https://doi.org/10.1002/dvg.22785

Dragomirka Jovic, c., Xue Liang, Hua Zeng, Lin Lin, Fengping Xu, and Yonglun Luocorresponding. (2022). Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine*, *12*. https://doi.org/10.1002/ctm2.694

Elena Dreosti, N. V. L., Matthias Carl, Emre Yaksi, Stephen W. Wilson (2014). Left-right asymmetry is required for the habenulae to respond to both visual and olfactory stimuli. *Current Biology*, *24*, 440-445. https://doi.org/https://doi.org/10.1016/j.cub.2014.01.016

Fabrizio Palumbo, B. S., Robbrecht Pelgrims, Emre Yaksi. (2020). Zebrafish dorsolateral habenula is required for updating learned behaviors. *Cell reports*, *32*. https://doi.org/https://doi.org/10.1016/j.celrep.2020.108054

Fore, S., Palumbo, F., Pelgrims, R., & Yaksi, E. (2018). Information processing in the vertebrate habenula. *Semin Cell Dev Biol*, *78*, 130-139. https://doi.org/10.1016/j.semcdb.2017.08.019

Greenacre, M., Groenen, P. J. F., & Hastie, T. e. a. (2022). Principal component analysis. *Nature Reviews Methods Primers 2*. https://doi.org/https://doi.org/10.1038/s43586-023-00209-y

Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., & Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*, *42*(2), 293-304. https://doi.org/10.1038/s41587-023-01767-y

Haque, A., Engel, J., Teichmann, S. A., & Lonnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med*, *9*(1), 75. https://doi.org/10.1186/s13073-017-0467-4

Hashikawa, Y., Hashikawa, K., Rossi, M. A., Basiri, M. L., Liu, Y., Johnston, N. L., Ahmad, O. R., & Stuber, G. D. (2020). Transcriptional and Spatial Resolution of Cell Types in the Mammalian Habenula. *Neuron*, *106*(5), 743-758 e745. https://doi.org/10.1016/j.neuron.2020.03.011

Hikosaka, O. (2010). The habenula: from stress evasion to value-based decision-making. *Nat Rev Neurosci*, *11*(7), 503-513. https://doi.org/10.1038/nrn2866

Hong, E., Santhakumar, K., Akitake, C. A., Ahn, S. J., Thisse, C., Thisse, B., Wyart, C., Mangin, J. M., & Halpern, M. E. (2013). Cholinergic left-right asymmetry in the habenulo-interpeduncular pathway. *Proc Natl Acad Sci U S A*, *110*(52), 21171-21176. https://doi.org/10.1073/pnas.1319566110

Jeong, I., Hansen, J. N., Wachten, D., & Jurisch-Yaksi, N. (2022). Measurement of ciliary beating and fluid flow in the zebrafish adult telencephalon. *STAR Protocols*, *3*(3), 101542. https://doi.org/10.1016/j.xpro.2022.101542

Jetti, S. K., Vendrell-Llopis, N., & Yaksi, E. (2014). Spontaneous activity governs olfactory representations in spatially organized habenular microcircuits. *Curr Biol*, *24*(4), 434-439. https://doi.org/10.1016/j.cub.2014.01.015

Jolliffe, I. T. (2002). *Principal Component Analysis, Second Edition* (2 ed.). Springer.

Jung, N., & Kim, T. K. (2023). Spatial transcriptomics in neuroscience. *Exp Mol Med*, *55*(10), 2105-2115. https://doi.org/10.1038/s12276-023-01093-y

Kenney, J. W., Steadman, P. E., Young, O., Shi, M. T., Polanco, M., Dubaishi, S., Covert, K., Mueller, T., & Frankland, P. W. (2021). A 3D adult zebrafish brain atlas (AZBA) for the digital age. *Elife*, *10*. https://doi.org/10.7554/eLife.69988

Lawson, R. P., Nord, C. L., Seymour, B., Thomas, D. L., Dayan, P., Pilling, S., & Roiser, J. P. (2017). Disrupted habenula function in major depression. *Mol Psychiatry*, *22*(2), 202-208. https://doi.org/10.1038/mp.2016.81

Lax, E., Friedman, A., Croitoru, O., Sudai, E., Ben-Moshe, H., Redlus, L., Sasson, E., Blumenfeld-Katzir, T., Assaf, Y., & Yadid, G. (2013). Neurodegeneration of lateral habenula efferent fibers after intermittent cocaine administration: implications for deep brain stimulation. *Neuropharmacology*, *75*, 246-254. https://doi.org/10.1016/j.neuropharm.2013.06.034

Lazaridis, I., Tzortzi, O., Weglage, M., Martin, A., Xuan, Y., Parent, M., Johansson, Y., Fuzik, J., Furth, D., Fenno, L. E., Ramakrishnan, C., Silberberg, G., Deisseroth, K., Carlen, M., & Meletis, K. (2019). A hypothalamus-habenula circuit controls aversion. *Mol Psychiatry*, *24*(9), 1351-1368. https://doi.org/10.1038/s41380-019-0369-5

Lecca, S., Meye, F. J., & Mameli, M. (2014). The lateral habenula in addiction and depression: an anatomical, synaptic and behavioral overview. *Eur J Neurosci*, *39*(7), 1170-1178. https://doi.org/10.1111/ejn.12480

Lee, E. J., Hong, S. K., Choi, D. H., Gum, S. I., Hwang, M. Y., Kim, D. S., Oh, J. W., & Lee, E. S. (2022). Three-dimensional visualization of cerebral blood vessels and neural changes in thick ischemic rat brain slices using tissue clearing. *Sci Rep*, *12*(1), 15897. https://doi.org/10.1038/s41598-022-19575-w

Lekk, I., Duboc, V., Faro, A., Nicolaou, S., Blader, P., & Wilson, S. W. (2019). Sox1a mediates the ability of the parapineal to impart habenular left-right asymmetry. *Elife*, *8*. https://doi.org/10.7554/eLife.47376

Lenz, M., Muller, F. J., Zenke, M., & Schuppert, A. (2016). Principal components analysis and the reported low intrinsic dimensionality of gene expression microarray data. *Sci Rep*, *6*, 25696. https://doi.org/10.1038/srep25696

Liu, C., Li, R., Li, Y., Lin, X., Zhao, K., Liu, Q., ... & Liu, L. . (2022). Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. *Developmental Cell*, *57*, 1284-1298. https://doi.org/https://doi.org/10.1016/j.devcel.2022.04.009

Luke Zappia, & Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience*, *7*, 1-9. https://doi.org/doi: 10.1093/gigascience/giy083

Masakazu Agetsuma, H. A., Tazu Aoki, Ryoko Nakayama, Mikako Takahoko, Midori Goto, Takayuki Sassa, Ryunosuke Amo, Toshiyuki Shiraki, Koichi Kawakami, Toshihiko Hosoya, Shin-ichi Higashijima & Hitoshi Okamoto (2010). The habenula is crucial for experience-dependent modification of fear responses in zebrafish. *Nature Neuroscience*, *13*, 1354–1356

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform ManifoldApproximation and Projection forDimension Reduction. *arXiv preprint arXiv:1802.03426*.

Mech, A. M., Merteroglu, M., Sealy, I. M., Teh, M. T., White, R. J., Havelange, W., Brennan, C. H., & Busch-Nentwich, E. M. (2021). Behavioral and Gene Regulatory Responses to Developmental Drug Exposures in Zebrafish. *Front Psychiatry*, *12*, 795175. https://doi.org/10.3389/fpsyt.2021.795175

Michel, L., Molina, P., & Mameli, M. (2024). The behavioral relevance of a modular organization in the lateral habenula. *Neuron*. https://doi.org/10.1016/j.neuron.2024.04.026

Morris, J. S., Smith, K. A., Cowen, P. J., Friston, K. J., & Dolan, R. J. . (1999). Covariation of activity in habenula and dorsal raphe nuclei following tryptophan depletion. *Neuroimage*, *10*. https://doi.org/https://doi.org/10.1006/nimg.1999.0455

Muller, J., Szklarczyk, D., Julien, P., Letunic, I., Roth, A., Kuhn, M., Powell, S., von Mering, C., Doerks, T., Jensen, L. J., & Bork, P. (2010). eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res*, *38*(Database issue), D190-195. https://doi.org/10.1093/nar/gkp951

Namboodiri, V. M. K., Rodriguez-Romaguera, J., & Stuber, G. D. (2016). The habenula. *Current Biology*, *26(19)*, R873-R877. https://doi.org/https://doi.org/10.1016/j.cub.2016.08.051

NCBI. (2024). *HEPACAM hepatic and glial cell adhesion molecule [Homo sapiens (human)]*. National Library of Medicine (US), National Center for Biotechnology Information;.

Norton, W. H. J. (2019). Screening for drugs to reduce aggression in zebrafish. *Neuropharmacology*, *156*, 107394. https://doi.org/10.1016/j.neuropharm.2018.10.023

Okamoto, H., Agetsuma, M., & Aizawa, H. (2012). Genetic dissection of the zebrafish habenula, a possible switching board for selection of behavioral strategy to cope with fear and anxiety. *Developmental neurobiology*, *72*, 386-394. https://doi.org/https://doi.org/10.1002/dneu.20913

Okamoto, H., & Aizawa, H. (2013). Fear and anxiety regulation by conserved affective circuits. *Neuron*, *78*(3), 411-413. https://doi.org/10.1016/j.neuron.2013.04.031

Pandey, S., Shekhar, K., Regev, A., & Schier, A. F. (2018). Comprehensive Identification and Spatial Mapping of Habenular Neuronal Types Using Single-Cell RNA-Seq. *Curr Biol*, *28*(7), 1052-1065 e1057. https://doi.org/10.1016/j.cub.2018.02.040

Percival P. D'Gama, T. Q., Mehmet Ilyas Cosacak, Dheeraj Rayamajhi, Ahsen Konac, Jan Niklas Hansen, Christa Ringers, Francisca Acuña-Hinrichsen, Subhra P. Hui, Emilie W. Olstad, Yan Ling Chong, Charlton Kang An Lim, Astha Gupta, Chee Peng Ng, Benedikt S. Nilges, Nachiket D. Kashikar, Dagmar Wachten, David Liebl, Kazu

Kikuchi, Caghan Kizil, Nathalie Jurisch-Yaksi. (2021). Diversity and function of motile ciliated cell types within ependymal lineages of the zebrafish brain. *Cell reports*, *37*. https://doi.org/https://doi.org/10.1016/j.celrep.2021.109775

Powell, G. T., Faro, A., Zhao, Y., Stickney, H., Novellasdemunt, L., Henriques, P., Gestri, G., Redhouse White, E., Ren, J., Lu, W., Young, R. M., Hawkins, T. A., Cavodeassi, F., Schwarz, Q., Dreosti, E., Raible, D. W., Li, V. S. W., Wright, G. J., Jones, E. Y., & Wilson, S. W. (2024). Cachd1 interacts with Wnt receptors and regulates neuronal asymmetry in the zebrafish brain. *Science*, *384*(6695), 573-579. https://doi.org/10.1126/science.ade6970

Proulx, C. D., Hikosaka, O., & Malinow, R. (2014). Reward processing by the lateral habenula in normal and depressive behaviors. *Nat Neurosci*, *17*(9), 1146-1152. https://doi.org/10.1038/nn.3779

Sakhi, K., Belle, M. D., Gossan, N., Delagrange, P., & Piggins, H. D. (2014). Daily variation in the electrophysiological activity of mouse medial habenula neurones. *J Physiol*, *592*(4), 587-603. https://doi.org/10.1113/jphysiol.2013.263319

Sartorius, A., Kiening, K. L., Kirsch, P., von Gall, C. C., Haberkorn, U., Unterberg, A. W., Henn, F. A., & Meyer-Lindenberg, A. (2010). Remission of major depression under deep brain stimulation of the lateral habenula in a therapy-refractory patient. *Biol Psychiatry*, *67*(2), e9-e11. https://doi.org/10.1016/j.biopsych.2009.08.027

Schmidt, E. R., & Pasterkamp, R. J. (2017). The molecular mechanisms controlling morphogenesis and wiring of the habenula. *Pharmacology Biochemistry and Behavior*, *162*, 29-37. https://doi.org/https://doi.org/10.1016/j.pbb.2017.08.008

Schmidt, F. M., Schindler, S., Adamidis, M. et al. . (2017). Habenula volume increases with disease severity in unmedicated major depressive disorder as revealed by 7T MRI. *Eur Arch Psychiatry Clin Neurosci 267*. https://doi.org/https://doi.org/10.1007/s00406-016-0675-8

Sharma, D., Worssam, M. D., Pedroza, A. J., Dalal, A. R., Alemany, H., Kim, H. J., ... & Quertermous, T. (2024). Comprehensive Integration of Multiple Single-Cell Transcriptomic Data Sets Defines Distinct Cell Populations and Their Phenotypic Changes in Murine Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *44*, 391-408. https://doi.org/https://doi.org/10.1161/ATVBAHA.123.320030

Shih, P. Y., Engle, S. E., Oh, G., Deshpande, P., Puskar, N. L., Lester, H. A., & Drenan, R. M. (2014). Differential expression and function of nicotinic acetylcholine receptors in subdivisions of medial habenula. *J Neurosci*, *34*(29), 9789-9802. https://doi.org/10.1523/JNEUROSCI.0476-14.2014

Sjostedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., Huang, J., Du, Y., Lin, L., Dong, Z., Yang, L., Liu, X., Jiang, H., Xu, X., Wang, J., . . . Mulder, J. (2020). An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*, *367*(6482). https://doi.org/10.1126/science.aay5947

Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods*, *18*(1), 100-106. https://doi.org/10.1038/s41592-020-01018-x

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888-1902 e1821. https://doi.org/10.1016/j.cell.2019.05.031

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, *6*(5), 377-382. https://doi.org/10.1038/nmeth.1315

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Res*, *25*(10), 1491-1498. https://doi.org/10.1101/gr.190595.115

Turner, K. J., Hawkins, T. A., Yanez, J., Anadon, R., Wilson, S. W., & Folgueira, M. (2016). Afferent Connectivity of the Zebrafish Habenulae. *Front Neural Circuits*, *10*, 30. https://doi.org/10.3389/fncir.2016.00030

Ulrike Hüsken, H. L. S., Gaia Gestri, Isaac H. Bianco, Ana Faro, Rodrigo M. Young, Myriam Roussigne, Thomas A. Hawkins, Carlo A. Beretta, Irena Brinkmann, Alessio Paolini, Raquel Jacinto, Shahad Albadri, Elena Dreosti, Matina Tsalavouta, Quenten Schwarz, Florencia Cavodeassi, Anukampa K. Barth, Lu Wen, Bo Zhang, Patrick Blader, Emre Yaksi, Lucia Poggi, Mihaela Zigman, Shuo Lin, Stephen W. Wilson, and Matthias Carl. (2014). Tcf7l2 Is Required for Left-Right Asymmetric Differentiation of Habenular Neurons. *Current Biology*, *24*, 2217-2227. https://doi.org/http://dx.doi.org/10.1016/j.cub.2014.08.006

Une, H., Yamasaki, R., Nagata, S., Yamaguchi, H., Nakamuta, Y., Indiasari, U. C., Cui, Y., Shinoda, K., Masaki, K., Gotz, M., & Kira, J. I. (2021). Brain gray matter astroglia-specific connexin 43 ablation attenuates spinal cord inflammatory demyelination. *J Neuroinflammation*, *18*(1), 126. https://doi.org/10.1186/s12974-021-02176-1

van de Haar, L. L., Riga, D., Boer, J. E., Garritsen, O., Adolfs, Y., Sieburgh, T. E., van Dijk, R. E., Watanabe, K., van Kronenburg, N. C. H., Broekhoven, M. H., Posthuma, D., Meye, F. J., Basak, O., & Pasterkamp, R. J. (2022). Molecular signatures and cellular diversity during mouse habenula development. *Cell Rep*, *40*(1), 111029. https://doi.org/10.1016/j.celrep.2022.111029

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*.

Viswanath, H., Carter, A. Q., Baldwin, P. R., Molfese, D. L., & Salas, R. (2013). The medial habenula: still neglected. *Front Hum Neurosci*, *7*, 931. https://doi.org/10.3389/fnhum.2013.00931

Wallace, M. L., Huang, K. W., Hochbaum, D., Hyun, M., Radeljic, G., & Sabatini, B. L. . (2020). Anatomical and single-cell transcriptional profiling of the murine habenular complex. *Elife*, *9*. https://doi.org/https://doi.org/10.7554/eLife.51271

Wang, Q., Ding, S. L., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer, B., Ho, A., Dolbeare, T., Blanchard, B., Dee, N., Wakeman, W., Hirokawa, K. E., Szafer, A., Sunkin, S. M., Oh, S. W., Bernard, A., . . . Ng, L. (2020). The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell*, *181*(4), 936-953 e920. https://doi.org/10.1016/j.cell.2020.04.007

Warden, M. R., Selimbeyoglu, A., Mirzabekov, J. J., Lo, M., Thompson, K. R., Kim, S. Y., ... & Deisseroth, K. (2012). A prefrontal cortex-brainstem neuronal projection that controls response to behavioural challenge. *Nature*, *492*, 428-432. https://doi.org/https://doi.org/10.1038/nature11617

Xie, B., Jiang, Q., Mora, A., & Li, X. (2021). Automatic cell type identification methods for single-cell RNA sequencing. *Comput Struct Biotechnol J*, *19*, 5874-5887. https://doi.org/10.1016/j.csbj.2021.10.027

Yamaguchi, T., Danjo, T., Pastan, I., Hikida, T., & Nakanishi, S. (2013). Distinct roles of segregated transmission of the septo-habenular pathway in anxiety and fear. *Neuron*, *78*(3), 537-544. https://doi.org/10.1016/j.neuron.2013.02.035

Yao, Z., van Velthoven, C. T. J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., Baker, K., Baker, P., Barkan, E., Bertagnolli, D., Bhandiwad, A., Bielstein, C., Bishwakarma, P., Campos, J., Carey, D., . . . Zeng, H. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature*, *624*(7991), 317-332. https://doi.org/10.1038/s41586-023-06812-z

Zeis, T., Allaman, I., Gentner, M., Schroder, K., Tschopp, J., Magistretti, P. J., & Schaeren-Wiemers, N. (2015). Metabolic gene expression changes in astrocytes in Multiple Sclerosis cerebral cortex are indicative of immune-mediated signaling. *Brain Behav Immun*, *48*, 313-325. https://doi.org/10.1016/j.bbi.2015.04.013

Zhang, M., Pan, X., Jung, W., Halpern, A. R., Eichhorn, S. W., Lei, Z., Cohen, L., Smith, K. A., Tasic, B., Yao, Z., Zeng, H., & Zhuang, X. (2023). Molecularly defined and

spatially resolved cell atlas of the whole mouse brain. *Nature*, *624*(7991), 343-354. https://doi.org/10.1038/s41586-023-06808-9

Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R., & Siebert, P. D. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques*, *30*(4), 892-897. https://doi.org/10.2144/01304pf02