UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Pan-brain Cell-type Profiling Using Fluorescence in situ Hybridization: How to Make a Periodic Table of Brain Cell Types?

Permalink

https://escholarship.org/uc/item/69j584tr

Author

Moghadami, Siavash

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Pan-brain Cell-type Profiling Using Fluorescence *in situ* Hybridization: *How to Make a Periodic Table of Brain Cell Types?*

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Chemistry

by

Siavash Moghadami

The committee in charge:

Professor Don W. Cleveland, Chair
Professor Wei Wang, Co-Chair
Professor Bogdan Bintu
Professor Kamil Godula

2022

Copyright

Siavash Moghadami, 2022

All rights reserved.

The Thesis of Siavash Moghadami is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
FISH	Fluorescence *in situ* Hybridization
smFISH	Single Molecule FISH
MERFISH	Multiplexed Error Robust FISH

ACKNOWLEDGEMENTS

I am standing on the shoulders of giants in my research process. First, I want to sincerely thank my principal investigators, Prof. Don W. Cleveland and Prof. Bogdan Bintu, for their continuous support and guidance; without their help, my educational journey would not be possible. Next, I want to thank my lab mates from Cleveland's lab who taught me how to do science: Dr. Roy Maimon, Dr. Dong Hyun Kim, Dr. Alexander Goginashvili, Dr. Shan Lu, Dr. Zevik Melamed, Carlos Marinas, Melissa McAlonis-Downes and my other lab mates from Bintu's lab who have been excellent friends and colleagues during my time at UCSD: Alex Monell, Kseniya Malukhina, Dr. Joy Zhou, and Dr. Van Nguyen.

Ultimately, I want to thank my wonderful fiancé. Without her, I wouldn't write these lines. I am forever in debt to her unconditional love.

Chapter 1 introduces the importance of making a periodic table of brain cell types and overviews various techniques for cell type classification.

Chapter 2 is the experimental design details and the methods used to perform the intended experiments.

Chapter 3 is the experimental results which contain unpublished materials that will be submitted for publication.

VITA

2021 Bachelor of Science in Biochemistry, University of California San Diego

2022 Master of Science in Chemistry: Chemical Biology, University of California San Diego

Chemistry

ABSTRACT OF THE THESIS

Pan-brain Cell-type Profiling Using Fluorescence *in situ* Hybridization:
How to Make a Periodic Table of Brain Cell Types?

by

Siavash Moghadami

Master of Science in Chemistry

University of California San Diego, 2022

Professor Don W. Cleveland, Chair
Professor Wei Wang, Co-Chair

Among all the organs, the brain has the most heterogeneous cellular composition, which mostly remained unclear. With recent advances in omics, the required technologies for unlocking the mysteries of brain heterogeneity are emerging. After reviewing different techniques, we strived to make a periodic table of brain cell types using in situ hybridization techniques using mice as model organisms.

CHAPTER 1: INTRODUCTION

Biological science has long been interested in converting one cell type to another. Fig. 1.1 demonstrates such a reaction.

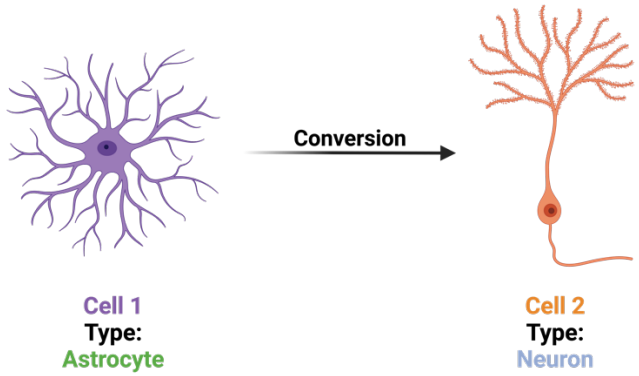

Fig. 1.1: Hypothetical reaction of converting a cell type to another.

Several pieces of information are needed to devise a successful method to perform this reaction scientifically. 1) What is the cell? 2) What is the cell type, and how to define it? 3) How to convert one cell type to another? I strive to answer these questions in the following sections.

1.1.What is the cell?

The question of "*what is life?*" has puzzled scientists throughout history. Before the invention of the light microscope, scientists didn't have a clear view of life. Life was considered as something descended from God. However, the notion of life has changed dramatically through reproducible experimentations after the invention of light microscopy.

In 1665, Robert Hooke published his landmark work titled "*Micrographia*" (Fig. 1.2), which was composed of his drawings, observations, and descriptions of the various forms of organisms under the bare light microscope [Hooke 1665]. Like many other examples in the history of science that developing new technologies and techniques resulted in discoveries, the invention of the light

microscope resulted in the discovery of the "cell" concept by Hooke for the first time while observing cork under the microscope. He coined the term "cell" due to the similarity of the pin box-shaped structures to rooms or cells of abbeys.

Hook's observations led to the classical Cell Theory proposed by Theodor Schwann and Mattias Schleiden in 1838 and 1839. According to the Cell Theory formulated in the mid-nineteenth century: 1) all living organisms are composed of at least one cell, 2) cells are basic units of life as well as structure and function in all known organisms, and 3) all cells are arising from pre-existing cells that have divided (proposed by Rudolf Virchow in 1858)

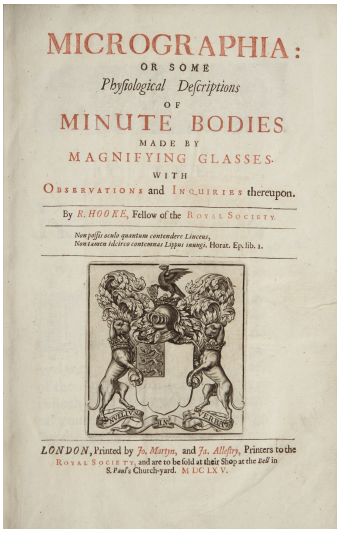
MICROGRAPHIA:
OR SOME
Physiological Descriptions
OF
MINUTE BODIES
MADE BY
MAGNIFYING GLASSES.
WITH
OBSERVATIONS and INQUIRIES thereupon.
By R. HOOKE, Fellow of the ROYAL SOCIETY.
Non possis oculo quantum contendere Lynceus,
Non tamen idcirco contemnas Lippus inungi. Horat. Ep. lib. 1.
LONDON, Printed by Jo. Martyn, and Ja. Allestry, Printers to the ROYAL SOCIETY, and are to be sold at their Shop at the Bell in S. Paul's Church-yard. M DC LX V.

Figure 1.2: The *Micrographia*.

With the advancement of various technologies, more detailed and controlled observations of the cells coming from all domains of life were made, which resulted in the formation of modern cell theory, which has three additional statements: 1) all cells contain hereditary genetic information encoded in the form of 2'-DeoxyriboNucleic Acids (DNA) that passed to offspring cells during cell division, 2) cells of the organisms from similar species are fundamentally similar

from both chemically and structurally perspectives, and 3) energy flows through the cells. Fig. 1.3 shows a generic animal cell and the associated organelles.

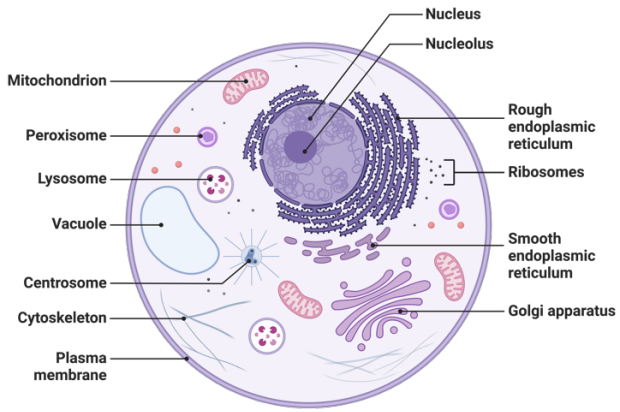

Figure 1.3: A generic animal cell diagram.

1.2.What is the Cell Type and How to Define Cell Type?

1.2.1. What is the Cell Type?

Although all cells have many similarities, cells come in different phenotypic features such as morphologies, sizes, and function. This cellular diversity contributes to the variety of known life within and between species. The formation of complex organisms with various cell types from a single-cell zygote involves the division, differentiation, and even apoptosis of cells or groups of cells throughout the organism's development. However, defining a cell type is not accessible due to various phenotypic characteristics at different levels. In humans only, it is estimated that there are more than 200 cell types with different morphologies specialized for specific functions. In a recent study using single-cell mRNA sequencing, more than 102 cell types were identified after sequencing 599,926 cells, as depicted in Fig. 1. 4 [Han 2020]. Not all organs have the same cellular heterogeneity. Among all the organs and all the organisms with the Central Nervous System (CNS), the brain has the most diverse cell types, as demonstrated in Fig. 1. 5 [Zeisel 2018].

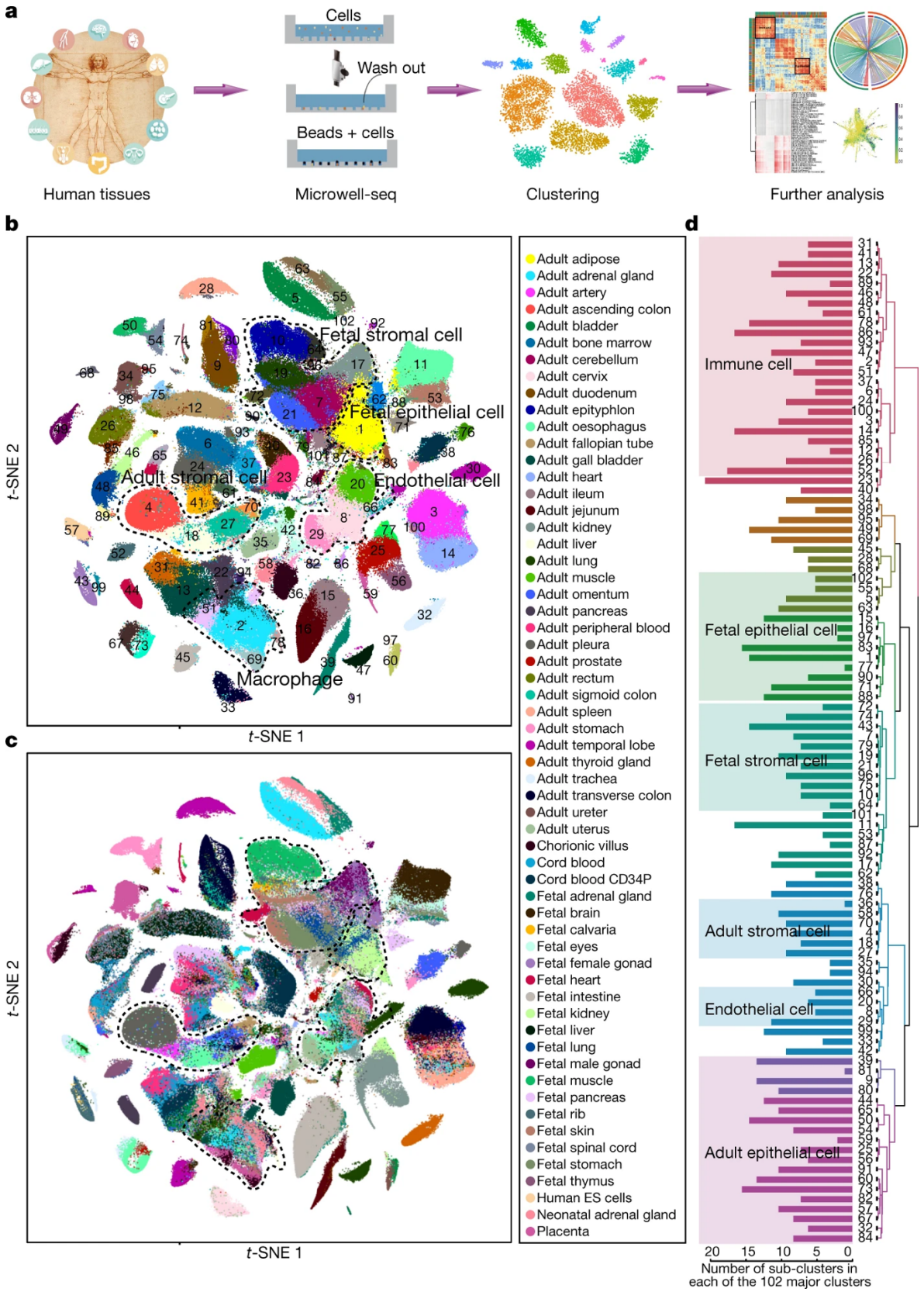

Figure 1.4: Diversity of human cell types using single-cell mRNA sequencing [Han 2020].

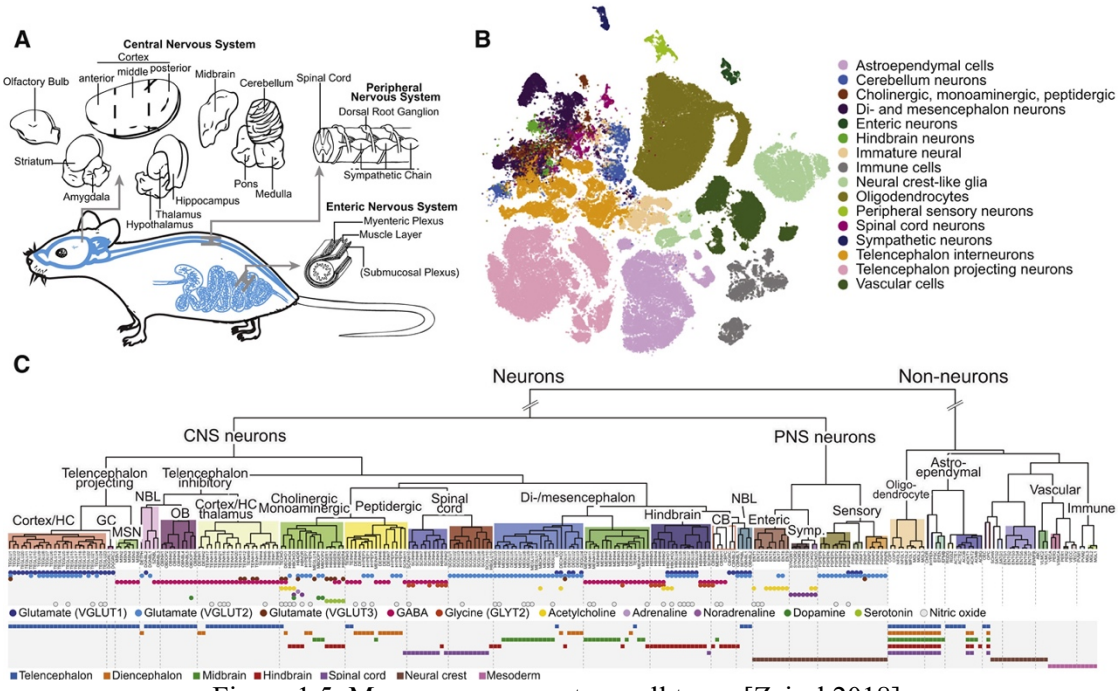

Figure 1.5: Mouse nervous system cell types [Zeisel 2018].

As a result, defining the cell type is crucial for better understanding the function of the cell and, subsequently, tissue within an organism.

Cells with similar structures and functions could be grouped to group the cells within an organism. Considering the presence of billions to trillions of cells within many organisms (multi-cellular life), cell type categorization enables the study of the cells more systematically and accurately by reducing the complexity of the system of interest. However, the main question is which features, properties or characteristics of the cells must be used to group and classify the cells. Unfortunately, a consensus on the exact cell type definition does not exist yet. Lack of cell type definition consensus not only makes the results of much research not accurately reproducible but also results in uncertainty in the investigation of all cell types in an organism. Despite the recent advances in transcriptomic techniques such as single cell/nucleus RNA sequencing, spatial transcriptomics, and *in situ* hybridization methods, it remains an unanswered question whether the grouped cells (clusters) represent biologically meaningful cell types or not. Moreover, although

cell type characterization may seem a relatively straightforward task, at the cellular level, individual cells exhibit heterogenous properties which not necessarily have high concordance with other similar cells within the same tissue making the cell type assignment a challenging task.

1.2.2. Cell Type Characterization Methods

Cells possess different phenotypic and functional features such as morphology, physiology, molecular, and location within and between tissues in multicellular organisms. Many studies tried categorizing cellular heterogeneity in the brain by addressing different cellular modalities [Fishell 2013; Mukamel 2019; Nelson 2006; Seung 2014; Zeng 2017]. However, no concrete consensus exists on how to classify the cells. As a result, unbiased quantitative methods with comprehensive coverage that yield reproducible results are needed to study all brain regions and then computationally classify the cells into different groups. Pursuing that completeness, only then can one say that all types of cells in the brain have been identified.

Various methods characterize cell types, as shown in Fig. 1.6 [Zeng 2022]. At the molecular level, RNA transcripts could be captured and sequenced to yield a cell transcriptome at a specific snapshot. Moreover, chromatin modifications such as accessible euchromatin, the 3D structure of the chromatin, or methylation of the chromatin could be profiled genome-wide (epigenome) to understand which parts of the chromatin are likely to be transcribed or related to each other. Also, although not popular so far, proteins of the single cells could be captured and identified using mass spectroscopy to have a proteome-level vision of the cell at that snapshot (Mund 2022; Cho 2022). Single-cell proteomics still needs a lot of technological advancements since there is not always a correlation between transcriptome and proteome. Moreover, if

developed with enough efficiency, spatial single-cell proteomics could dramatically give a better view of the distribution of proteins throughout the cells and tissues.

Considering a cell as a drawing, all these layers of information (genome, transcriptome, epigenome, metabolome, and proteome) are different layers of the final cell, portraying necessary to have a complete understanding of the whole cell.

Combining the molecular information of the single cells with their spatial information would result in a better understanding of individual cells in the context of the tissue and the connection between cells or connectomes. Usually, spatial transcriptomic techniques are based on *in situ* hybridization and imaging, *in situ* capture, or *situ* sequencings [Close 2021; Larsson 2021; Moses 2022; Zhuang 2021; Moffitt 2018]. Currently, single-cell or single-nucleus RNA sequencing is the most widely used technique to define the cell type due to its unbiased nature, comprehensiveness, and scalability to thousands of cells. There have been many attempts to make a cell atlas for different organisms [Lindeboom 2021; Regev 2017] or brain-specific [BRAIN Initiative Cell Census Network (BICCN) 2021; Ngai 2022].

Although transcriptome profiling methods have been so popular lately, cellular morphology and connectivity are critical defining features of brain cells. Many techniques based on Light Microscopy (LM) (high resolution or super-resolution microscopy) have been developed to profile the morphology of the cells in the brain [Gao 2022]. Besides LM, Electron Microscopy (EM) techniques have been developed to investigate the connectome and projection of the brain [Kulse 2021]. More technology development is needed in the connectivity of neurons at the level of cellular junctions since synapses are basic units of information processing in the brain. Without knowing the nature of the synapses, whether they are excitatory or inhibitory, and how they achieve their specificity, understanding or predicting the behavior of the whole neural circuit

would not be possible [Südhof 2021]. Recent attempts have been made to address this issue using super-resolution microscopy, but more technological advancement is needed to include synapses in the brain's connectome [Trotter 2019].

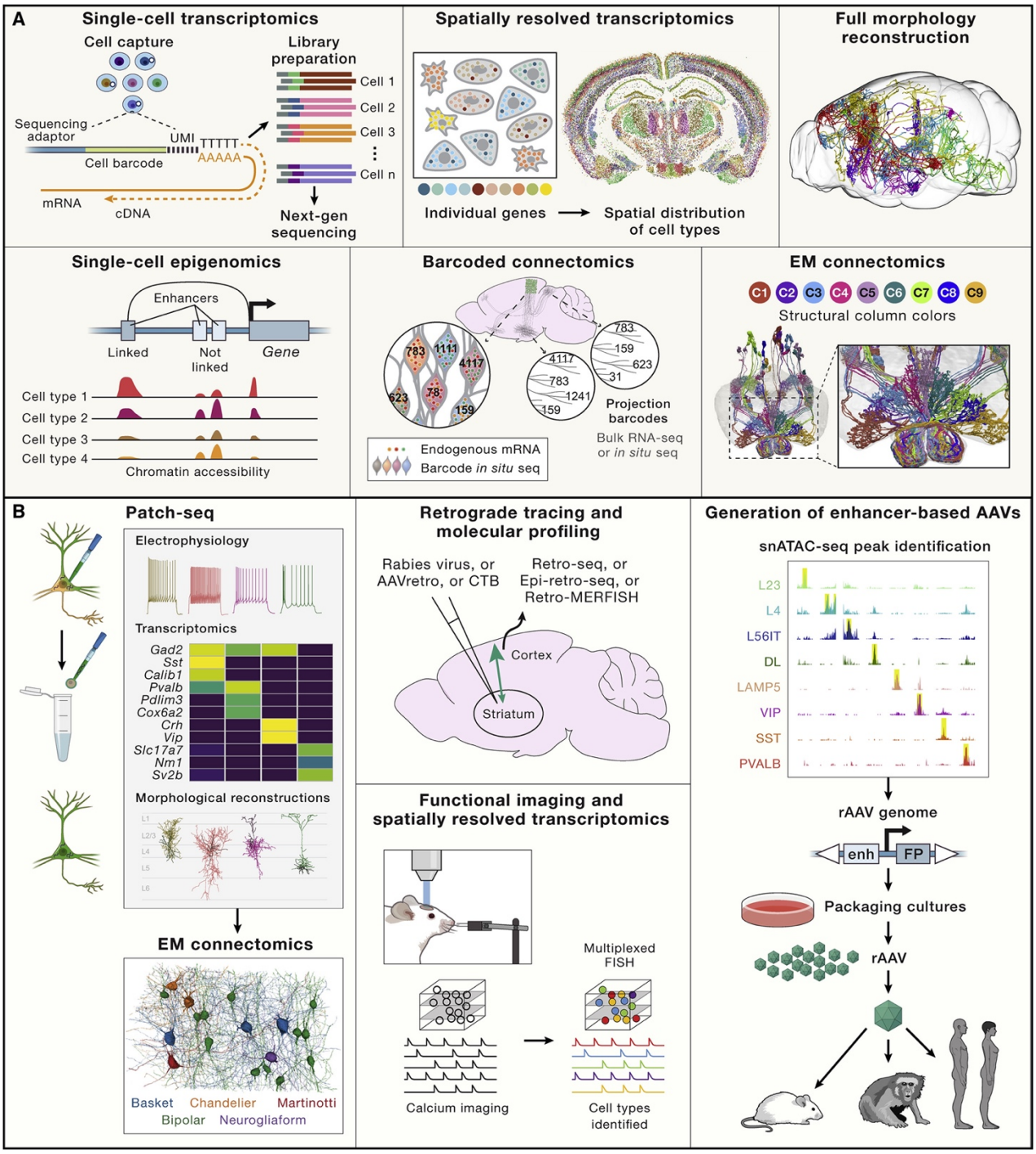

Fig. 1.6: Cell type characterization methods [Zeng 2022].

Although single modality measurements of single cells have yielded valuable insights, integrating multiple modality measurements of the same cell is critical to having a better and more coherent vision of the cell at different levels. Currently, the most common integration method is combining the cell's transcriptome with one or more modalities, as demonstrated in Fig. 1.6. There are many approaches for integrating different modalities. One possible approach is performing multimodal measurements of the same cell. Many methods have been developed considering this approach, such as single-cell muti-modal omics [Chenxu 2020], Patch-seq, which combines transcriptome and electrophysiology of the same cell [Berg 2021; Scala 2021; Le Floch 2022], and chemical connectivity tracing, such as Retro-seq, Epi-retro-seq, and Retro-MERFISH [Kim 2020; Tasic 2018; Zhang 2021], and calcium-based *in vivo* imaging combined with fluorescent *in situ* hybridization (FISH) [Bugeon 2022; Condylis 2022; Lovett-Barron 2020; Xu 2020].

Besides the mentioned methods, there are computationally based approaches to integrate different modalities, such as transferring labels across individually performed measurements. In this type of integration, critical features of other modalities are utilized for integration, such as marker genes of cellular transcriptome and accessible chromatin regions or chromatin modifications [Yao 2021]. Moreover, using all the information gathered in previous techniques, various cell-type-specific genetic tools could be developed to collect further information regarding a cell type [Daigle 2018; Matho 2021].

As with organisms, cell type diversification is also subjected to evolution. A new set of unique regulatory mechanisms and signatures (cell type-specific core regulatory complex of transcription factors) is required to establish a new cell type. This new set of regulatory instructions dictates the transcriptome of the cells and defines the cellular type. As a result, the cell type identity is written in its genome because diverse cell types are products of evolution. Consequently, the

gene expression pattern of a cell (transcriptomic and epigenomic) is an acceptable feature for cell type classification.

1.3.How to Convert a Cell Type to Another?

As shown in Fig. 1.7, until the late 19th century, it was thought that as cells differentiate, they only keep the genetic information necessary to preserve their unique identity and functions. Therefore, differentiation was widely believed to be an irreversible process [Morris 2019].

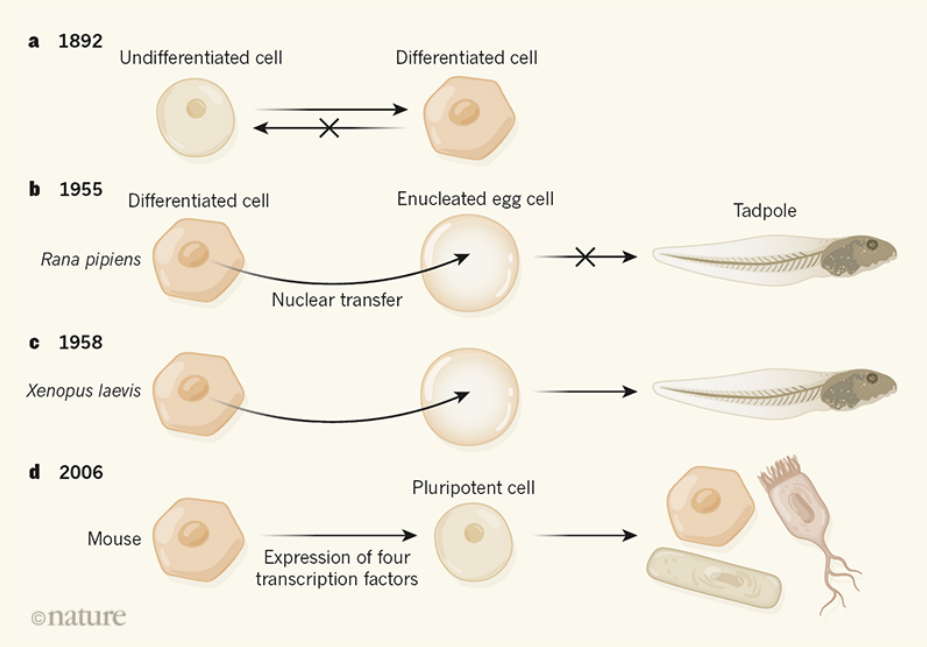

Fig. 1.7: Chronological advancement of the cellular differentiation process [Morris 2019].

In a study published in 1955 [Briggs 2019], the researchers transferred the nuclei of a blastula or differentiated cell to an enucleated egg. Although they could make a viable frog from blastula nuclei, no offspring were observed from differentiated nuclei. As a result, they concluded that the nuclei of cells in late-stage gastrula have an intrinsic restriction in differentiation potential. This viewpoint has been altered by further experiments [Gurdon 1958]. Briefly, scientists used another frog species in these experiments to transfer the nuclei of a differentiated cell into an

enucleated egg and produce a normal viable animal that could swim. This experiment demonstrated that cells retain their genomes intact, and no part of the genetic materials has been lost throughout the differentiation process.

Consequently, these results seriously challenged widely held dogma and opened a new perspective on the cell reprogramming field. Eventually, researchers discovered four well-known transcription factors to reprogram a mouse fibroblast cell into a pluripotent stem cell capable of differentiating into other cells. The demonstrated reprogramming process demonstrated that not only is the cell differentiation process reversible, but so is cellular identity [Takahashi 2006]. Moreover, these experiments could also conclude that the cell transcriptome or gene expression profile determines cellular identity supporting previous discussions.

Cellular reprogramming is very appealing in many cases, such as neurological disorders with the hallmark of neuronal death. The process would be highly beneficial therapeutically if a method could be developed to reprogram non-neuronal cells *in vivo* to replace lost neurons. As shown in Fig. 1.8, Different approaches to reprogramming a cell type to another are possible, including direct reprogramming or transdifferentiating and in-direct reprogramming. Each of these approaches has advantages and disadvantages. In direct reprogramming, cells would not go through a pluripotency/multipotency state, resulting in a brief therapy effect timewise.

Moreover, transdifferentiating cells would preserve the epigenomic features of the cells and decrease the chance of tumorigenesis due to a lack of pluripotency state. They would be suitable for *in vivo* cellular regeneration, specifically in the brain. In contrast to direct reprogramming, in in-direct cellular reprogramming, cells would go through the intermediate pluripotent/multipotent stem cells, resulting in a slow therapy effect.

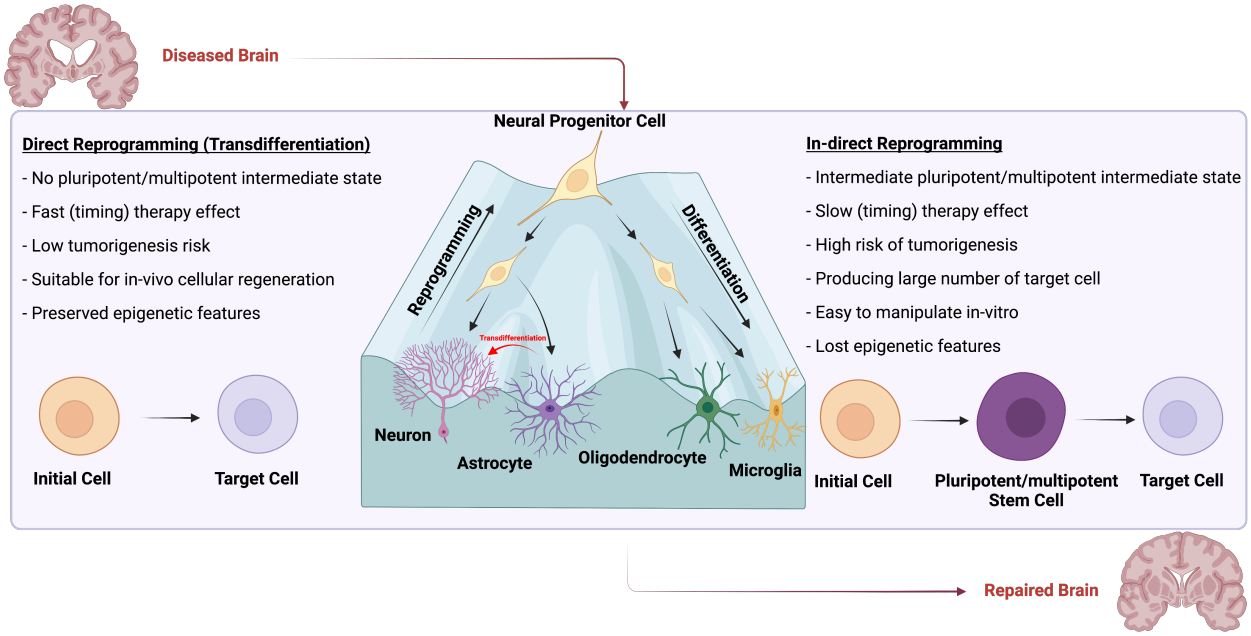

Fig. 1.8: Possible cellular reprogramming approaches.

Moreover, due to passing the pluripotency state, the epigenetic features of the cells would be lost, and the risk of tumorigenesis would be high. However, many target cell types could be generated. The advantage of in-direct reprogramming could be the possibility of manipulating and engineering the cells *in vitro*.

Regardless of the approach, to have a successful transformation, the conversion process should be tightly regulated by defining the cell type of the initial cells that will be converted and the target terminal cell (in this thesis, neurons) type. However, as mentioned before, no consensus exists on all brain cell types. As a result, in this thesis, we strived to generate a periodic table of the brain's cell types using the fluorescence *in situ* hybridization technique (Fig. 1.9) to lay a solid foundation for a successful neural cell reprogramming process which will be discussed in next chapter in detail.

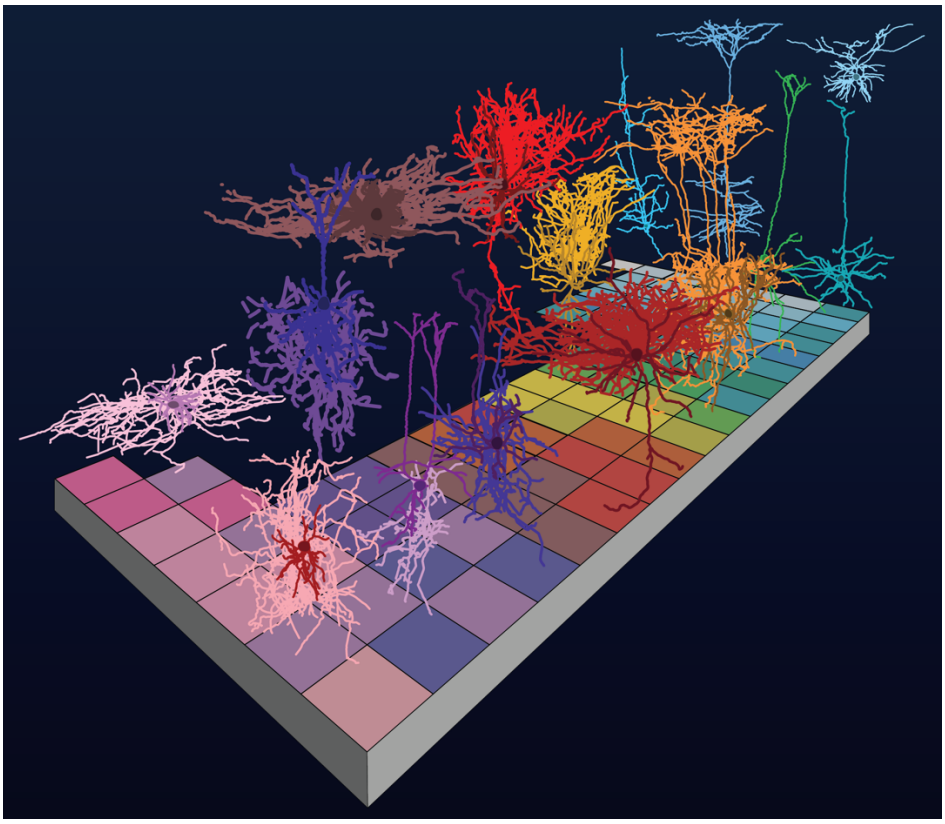

Fig. 1.9: Periodic table of the human brain [adopted from Allen Institute for Brain Science, Image credit to *Benedicte Rossi*].

CHAPTER 2: EXPERIMENTAL DESIGNS AND METHODS

There are several possible approaches for defining the cell type in a brain-wide fashion and completing the periodic table of the brain cell types. One possible method is using conventional single cell/nucleus RNA sequencing to define the transcriptome of the cells in different brain regions and classify them based on their gene expression pattern. Although, as mentioned before, transcriptome could be a suitable candidate for cell type classification, the spatial organization of the cells would be lost due to the process of preparing the samples for the sequencing process. Spatial organization of the cells across the brain is a crucial metric in classifying the cell type-specific in a brain-wide fashion. Another possible approach is using multi-modalities single cell/nucleus sequencing approaches to investigate both the transcriptome and epigenome of the cells and then classify the cells based on their transcriptome and epigenome. In this approach, although the efficiency of detecting transcripts is compromised due to adding another modality, having epigenomic information of the same cell could be very beneficial. However, again, spatial information of the cells would be lost due to the sequencing process. As a result, we decided to use Multiplexed Error Robust Fluorescence *in situ* Hybridization (MERFISH) to capture both the transcriptome and spatial information of the cells across the mice brain to make an atlas of the cell types in a brain-wide fashion.

2.1.From FISH to MERFISH

As shown in Fig. 2.1, Fluorescence *in situ* Hybridization (FISH) is a technique that individual molecules can be imaged with sub-cellular or single-molecule resolution and visualized while preserving spatial information. In conventional FISH for RNA imaging which is also known as single-molecule FISH (smFISH), complementary strands of the target RNAs are synthetically

synthesized, and the tissue is fixed on a slide. Then, the tissue is permeabilized, and synthesized RNA probes conjugated to fluorophores with different colors are introduced to the system. The designed probes would then be hybridized and annealed to the target RNA molecules via the canonical Watson-Crick base pairing process. Then using imaging techniques, individual molecules could be visualized.

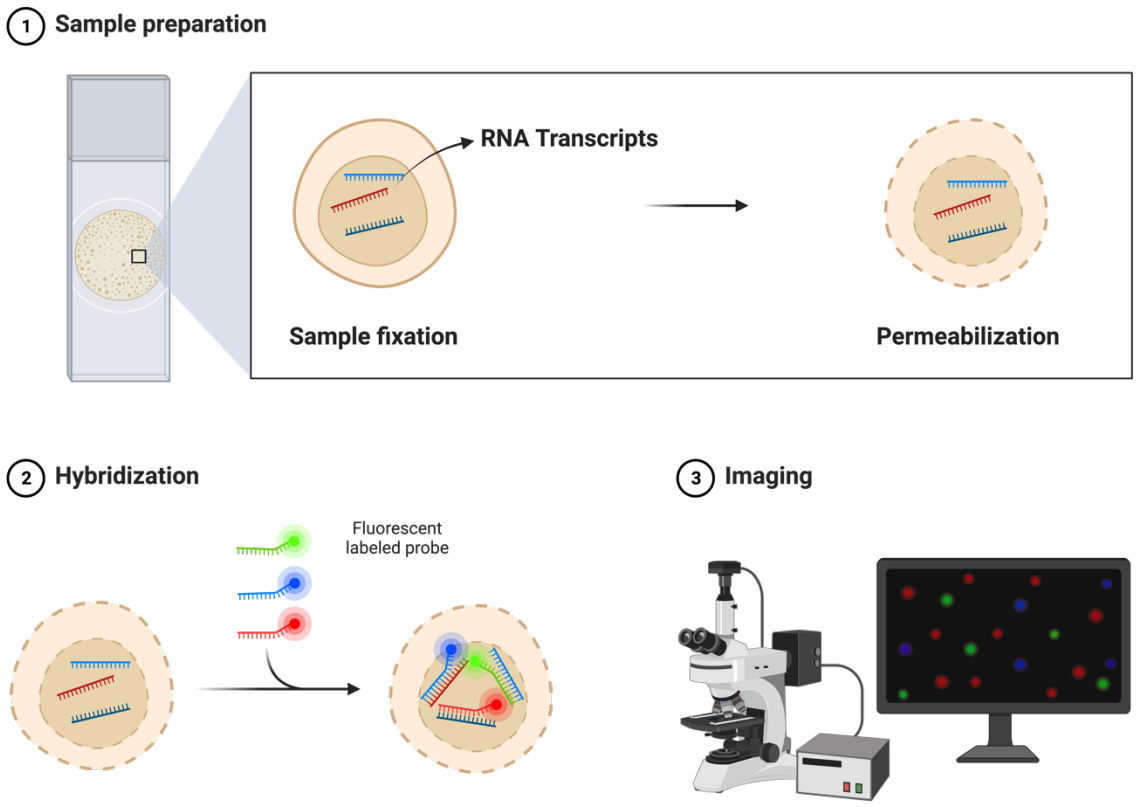

Fig. 2.1: Fluorescence *in situ* Hybridization (FISH)

The drawback of using FISH for determining the transcriptome of the cells is the limited number of fluorophores that could be conjugated to the probes (usually a maximum of four colors). Consequently, if many genes are required to be imaged, which is the case in cell type profiling, sequentially reflecting the probes is not a suitable approach. A multiplexed system is needed to

decrease the number of rounds of hybridization, imaging time, and data storage size. Fortunately, a Multiplexed FISH approach has been demonstrated recently [Chen 2015] (Fig. 2.2).

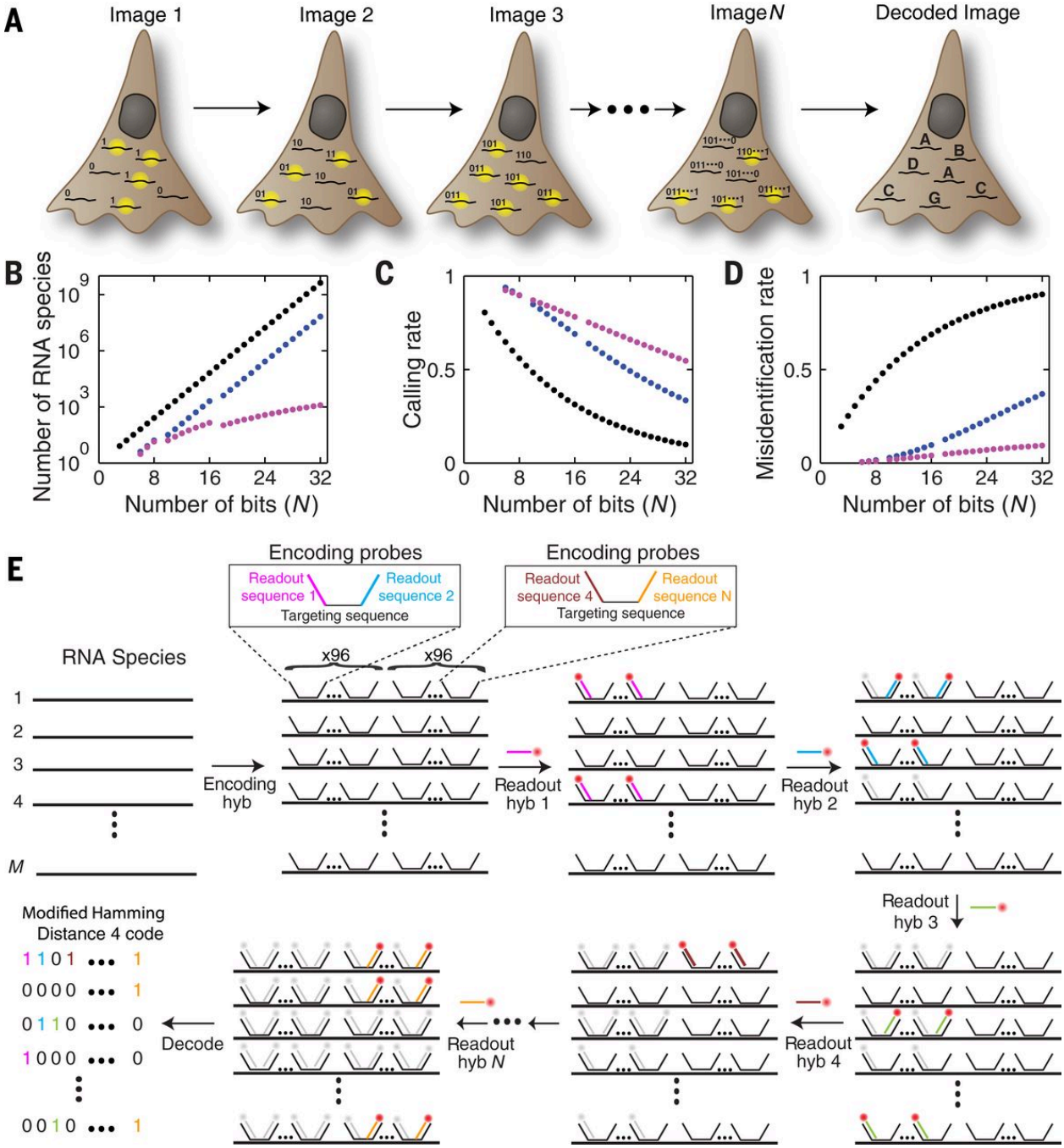

Fig. 2.2: MERFISH technology [Chen 2015].

In MERFISH, each RNA transcript is labeled with a certain number of complementary probes (target sequence) in different locations of the target RNA. By annealing these target

sequences to the RNA transcript, the RNA molecule is transformed into a specific combination of the readout sequences (coding strategy). Each probe has a central RNA-targeting region flanked by two distinct readout sequences chosen from a set of N particular sequences. Each of these N unique sequences is associated with a specific hybridization round. The coding strategy for each type of RNA is composed of a different combination of four N readout sequences (on bits) associated with the specific four rounds of hybridization that the target RNA transcript should have a read of 1.

Consequently, to probe the readout sequences, N rounds of hybridization with fluorescently conjugated readout probes are used (named Hyb1 to HybN). To minimize the error, the photobleaching process is applied after each round of hybridization to photobleach and inactivate the bound probes. There are several considerations in the design process of a MERFISH experiment. First, the target RNA transcripts must be known since the probes must be synthesized to target them. Consequently, the measured transcriptome might be a biased and unbiased survey of gene expression that is not feasible. The encoding strategy should be designed carefully to prevent optical crowding and reach peak efficiency.

2.2.Determining Target RNAs for Brain-wide Cell Type Profiling

As mentioned in the previous section, the first step of designing a MERFISH experiment is choosing the target RNA transcripts that will be imaged. Several pan-brain or semi-pan-brain single-cell/nucleus RNA-seq datasets (Table 2.1) have been obtained to select the target genes. We tried using datasets not enriched for a specific cell type, although not all the datasets in Table 2.1 are fully pan-brain. Unfortunately, there are not that many pan-brain high-quality datasets

available. However, many recent attempts have been made to address this challenge using various techniques.

Table 2.1: Pan-brain single-cell/nucleus RNA-seq datasets.

Dataset Number	Type (Cell/Nucleus) Count	Year	Tissue Technology/Time Points	Reference
1	Nucleus 469.2k	2019	Brain-wide 10XV2 (P60)	Not Published
2	Cell 6.5M	2021	Primary Motor Cortex 10XV3	Yao 2021
3 4 5	Cell 1.3M 2.5M 157.4k	2021	Mostly Neurons 10XV3 10Xv2 SMART-seq V4	Tasic 2018 Yao 2021
6 7	Nucleus 90.9k (V2) 90.4k (V3)	2021	Mostly Neurons 10XV2 10XV3	
8	Cell 98.0k	2021	Neocortex 10XV2 (E12.5 to E18.5 P1 and P4 (Multi-ome)	J Di Bella 2021
9	Cell 5.5k	2018	Dentate Gyrus 10XV1 (P12, P16, P24, P35)	Hochgerner 2018
10	Cell 509.9k	2018	19 Brain Regions 10XV1,10XV2 (P20-P30)	Zeisel 2018
11	Cell 690k	2018	9 Brain Regions Drop-seq (P60-P70)	Saunders 2018
12	Cell 1.3M	2020	Brain-wide 10Xv2 (E18)	10X Sample Dataset
13	Cell 1.1M	2021	Cortex and Hippocampus 10Xv2, SSv4	Yao 2021
14	Cell 292.5k	2021	Brain-wide 10Xv1 (Gastrulation to Birth, Time Points)	La Manno 2021

For example, we used dataset 10 [Zeisel 2018], which classified cell types across the adult mice nervous system (CNS and PNS) using single-cell RNA-seq on ~510k cells taken from 19 different regions of the adult mice nervous system, as shown in Fig. 2.3.

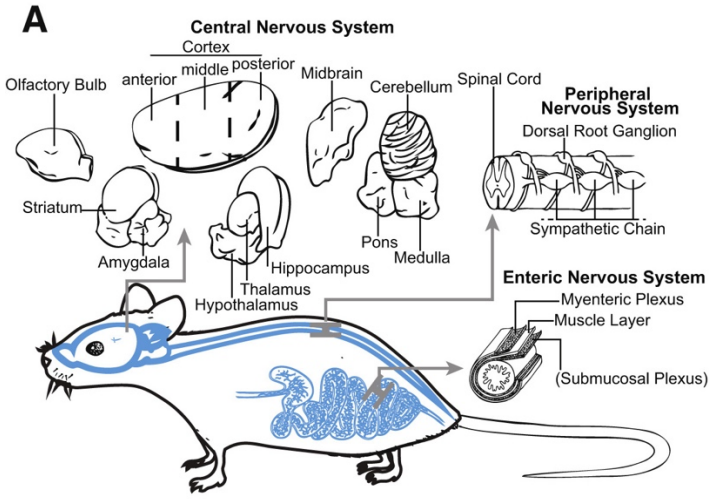

Fig. 2.3: 19 mice's nervous system regions profiled in [Zeisel 2018].

Next, as shown in Fig. 2.4, we first clustered the cells using the Leiden clustering method at various hierarchical levels. At each clustering level, we applied a machine learning algorithm called Necessary and Sufficient Forest v3.0 (NS-Forest v3) described in [Aevermann 2021], which uses random forests to find a small subset of the transcriptome that holds enough differential expression information to preserve the integrity of the cell clustering at each clustering level. Then, we refined the output genes using biological knowledge and added additional genes such as proliferation markers and housekeeping genes such as *Malat1*. The NS-Forest algorithm iterates over each cell type in the dataset. It constructs a Random Forest classifier to separate the target cell type from all other cell types present in the dataset at that level of clustering. Then, this algorithm identifies 15 genes with the highest Gini index for the random forest. From these 15 genes, we selected six genes with the most binary expression, primarily expressed in only the target

cell type, to be evaluated in combination to maximize the Fbeta score, where beta was set to 0.5. The algorithm's output would be a group of cell type markers with high precision in classifying a given cell into its actual cell type, rather than a higher beta value which would sacrifice accuracy for a more extensive set of markers. Fig. 2.4 shows the Leiden clustering of the [Zeisel 2018] in the first clustering level.

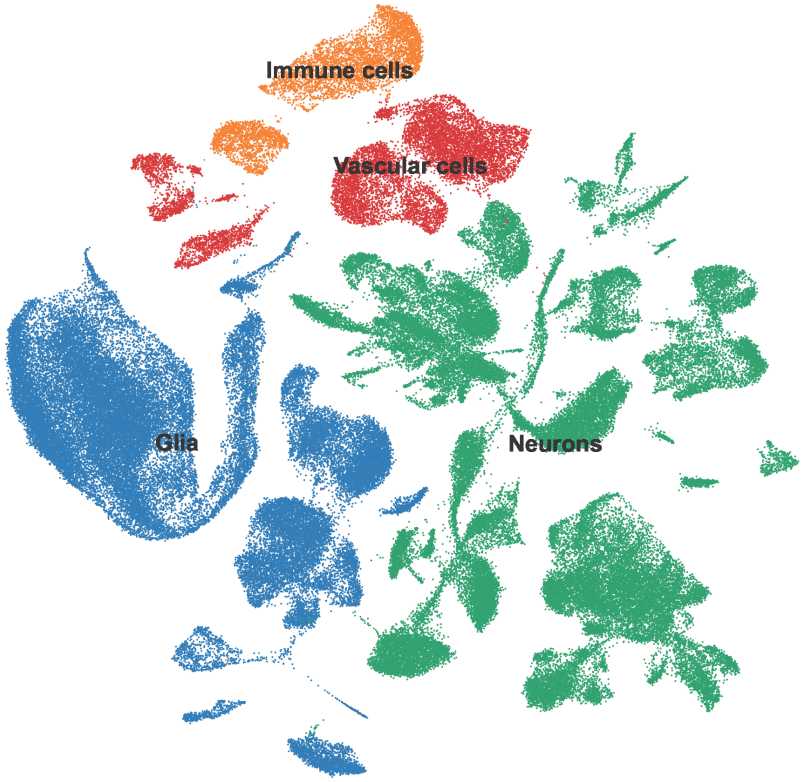

Fig. 2.4: Clustering of [Zeisel 2018] at clustering level 1.

At this level of clustering, we tried to include target genes that could be used for high-level characterization of the cell types across the brain, such as neurons, glia, and vascular cells. Fig. 2.5 demonstrates the next level of clustering, which is more comprehensive than level 1. At this level of clustering, more detailed yet high-level cell clusters could be identified, and the unique marker genes of each cluster were included in the final set of target genes. We excluded the cells from PNS regions since we focus only on the brain.

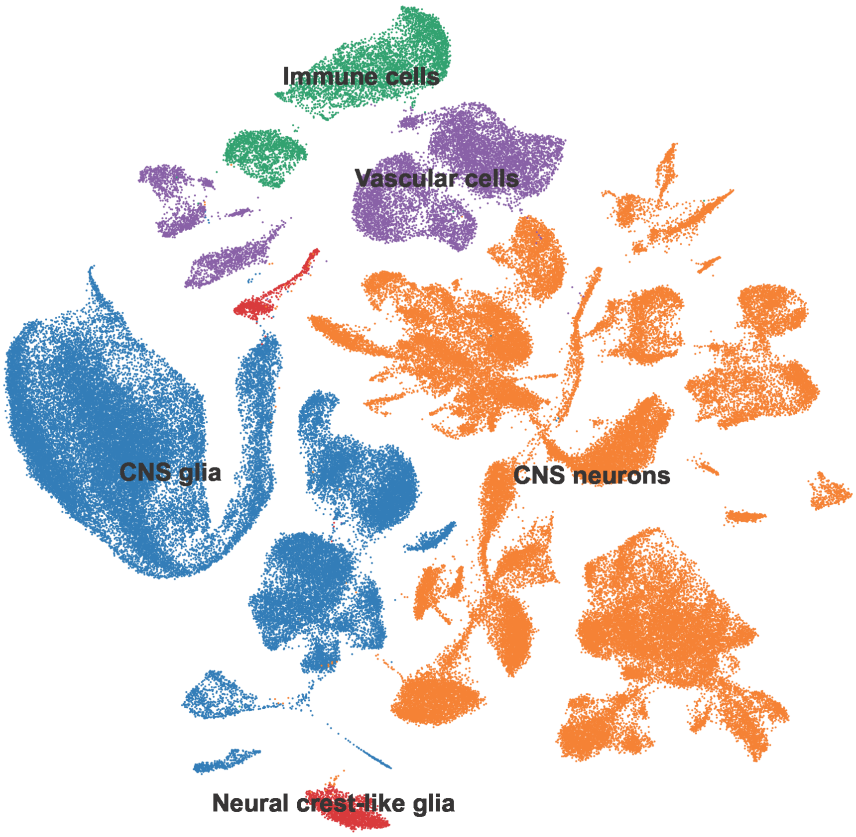

Fig. 2.5: Clustering of [Zeisel 2018] at clustering level 2.

Fig. 2.6 and Fig. 2.7 demonstrate the clustering of the same dataset at level 3 and level 4, respectively. We tried to make sure that at each level (from level 1 to level 4), the same cell clusters were not breakdown into different subtypes to make sure that we had captured enough target genes to be able to distinguish significant cell types from each other before going into subtypes of each primary cell type to increase our accuracy in the MERFISH pan-brain dataset.

The periodic table of brain cell types would not be complete if only major cell types such as neurons, astrocytes, oligodendrocytes, and microglia could be distinguished. To capture subtypes of primary cell types, we have applied NS-Forest v3 to the same dataset at an even finer resolution, as shown in Fig. 2.8. By doing so, we made sure that we can distinguish different

subtypes of neurons, astrocytes, and other types of cells at various regions of the brain which is crucial for cell type profiling of the whole brain.

Fig. 2.6: Clustering of [Zeisel 2018] at clustering level 3.

Each subtype of a primary cell type in the brain should be distinguished from the other since cells from each subtype would have different molecular and functional characteristics in the brain. For example, it has been reported that several different subtypes of astrocytes exist in the nervous system [Liddelow 2017]. Moreover, the role of cellular subtypes would be crucial in reprogramming one cell type to another. For example, recently, it has been reported that by transient downregulation of Ptbp1 in the mouse brain, mature astrocytes could be converted to

mature neurons [Qian 2020]. However, if transdifferentiating, it is still unclear what starting cells are converting to mature neurons. Our approach could be coupled with such experiments to better characterize the origin of the converting and terminal cells in a pan-brain fashion.

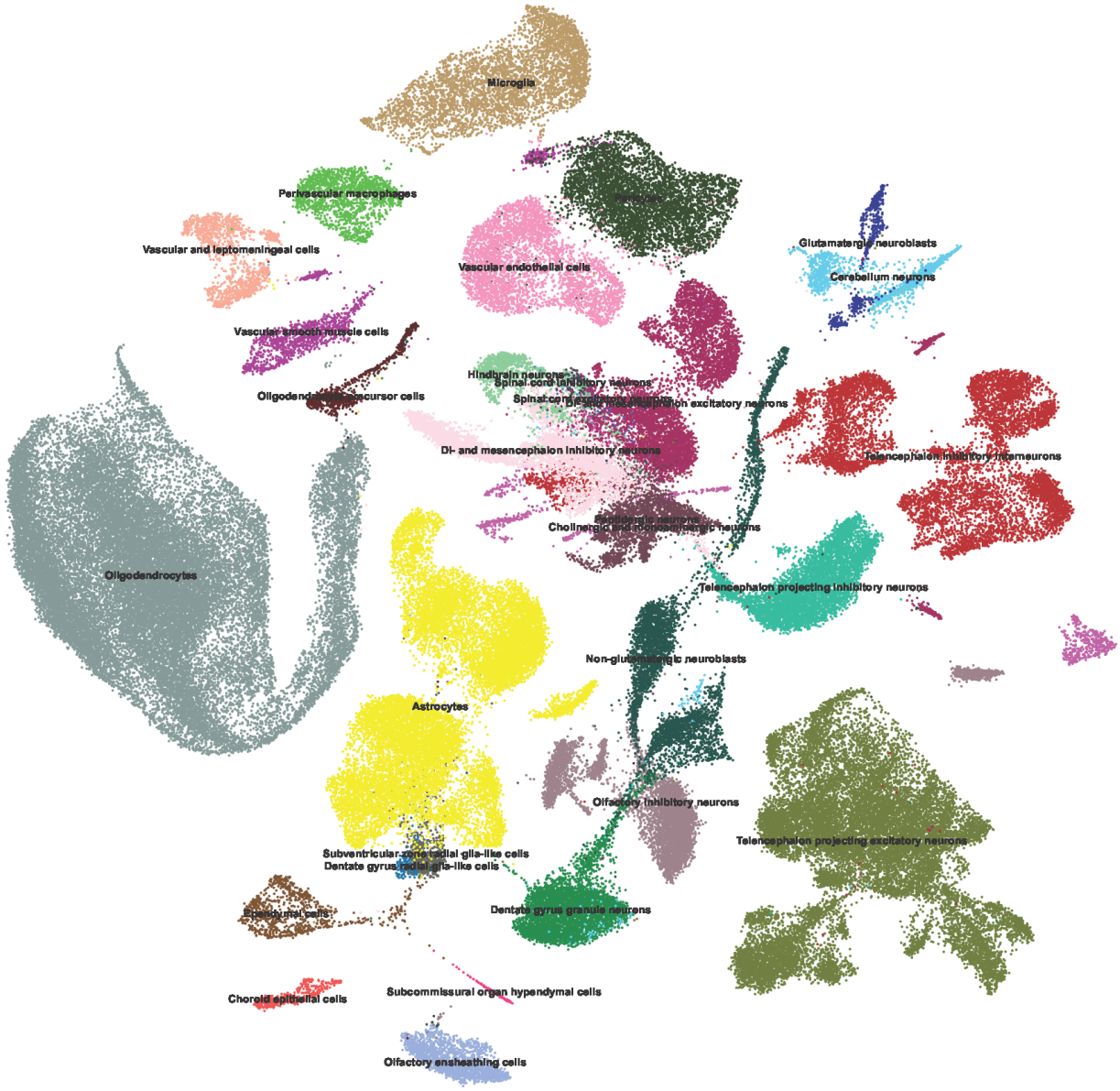

Fig. 2.7: Clustering of [Zeisel 2018] at clustering level 4.

By analyzing the [Zeisel 2018] dataset, we obtain 331 target genes for profiling the cell types across the brain. In comparison to the initial clustering of the cells with 3048 highly variable

genes, we reduced the target genes from 3048 to 331 while preserving the clustering accuracy, which shows the success of our approach.

Fig. 2.8: Clustering of [Zeisel 2018] at sub-major cell type clustering.

Fig. 2.9 compares the clustering accuracy while clustering with 3048 highly variable genes with our identified 331 genes. As shown, the overall shape of the clusters was well preserved, and all the major cell types are distinguishable. We applied the same approach to all other datasets,

and after each analysis, we included the unique target genes that were not identified in the previous study. As was expected, as more datasets were analyzed, fewer and fewer unique target genes were identifiable.

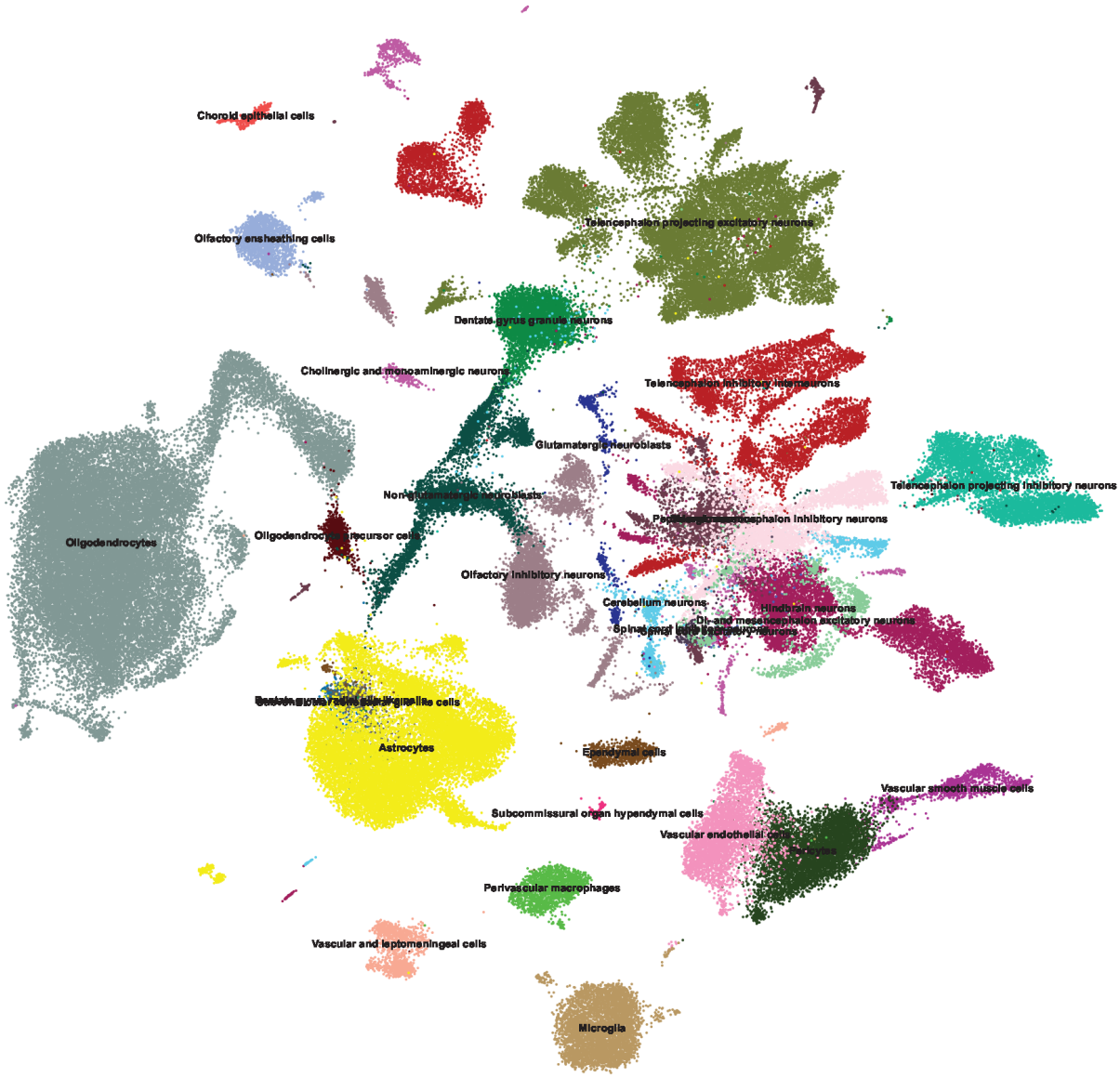

Fig. 2.9: Clustering of [Zeisel 2018] with identified 331 target genes.

After analyzing all the datasets, we identified 585 target genes necessary for cell type profiling pan-brain. We also included cell cycle markers, highly expressed housekeeping genes,

genes related to the neural transdifferentiation process that we are doing in parallel to this project, and marker genes from literature and synthesized a library of ~1500genes.

After synthesizing our list of genes, we strived to quantitatively test our target genes' accuracy in cell type identification. We assigned a new cell type label to each cell in the test dataset to calculate the clustering accuracy. This new cell type label represents the most prevalent cell type of the five nearest neighbors in terms of gene expression to the cell. These five nearest cell neighbors are found by calculating a Euclidean distance matrix between gene expression profiles of all the cells. Once each cell is reassigned to a cluster, the accuracy of the clustering has been determined by dividing the number of cells reassigned to their original cluster by the number of total cells reassigned. The result is reported in Table 2.2. As shown in this table, with our targeted genes, we could accurately identify the cell types with high accuracy even compared to the list of well-established marker genes that were reported in [Zeisel 2018]. Moreover, our target genes perform better in comparison to only using top differentially expressed genes of all clusters. Our 680 target genes were able to separate cell types among the pan-brain datasets listed in Table 2.2 with clustering accuracies exceeding those of similarly sized gene sets and close to the large ground truth high variable gene sets.

Table 2.2: Comparison of clustering accuracy.

Target Genes	Finest Cell Type Accuracy [Zeisel 2018]	General Cell Type Accuracy [Zeisel 2018]	Allen Brain General Cell Type Accuracy [Yao 2021]	Allen Brain Finest Cell Type Accuracy [Yao 2021]
Our 680 Target Genes	85%	98%	99%	85%
819 Differentially Expressed Genes of [Zeisel 2018]	80%	97%	99%	83%
3048 Highly Variable Genes of [Zeisel 2018]	86%	99%	99%	87%
440 Reported Marker Genes of [Zeisel 2018]	76%	98%	99%	84%
2344 Highly Variable Genes of [Yao 2021]	87%	99%	99%	90%

2.3.Designing the Encoding Strategy and Probes

The next step after picking the target genes for imaging is devising an encoding strategy and designing the probes accordingly. Fig. 2.10 shows our process for encoding strategy and probe design.

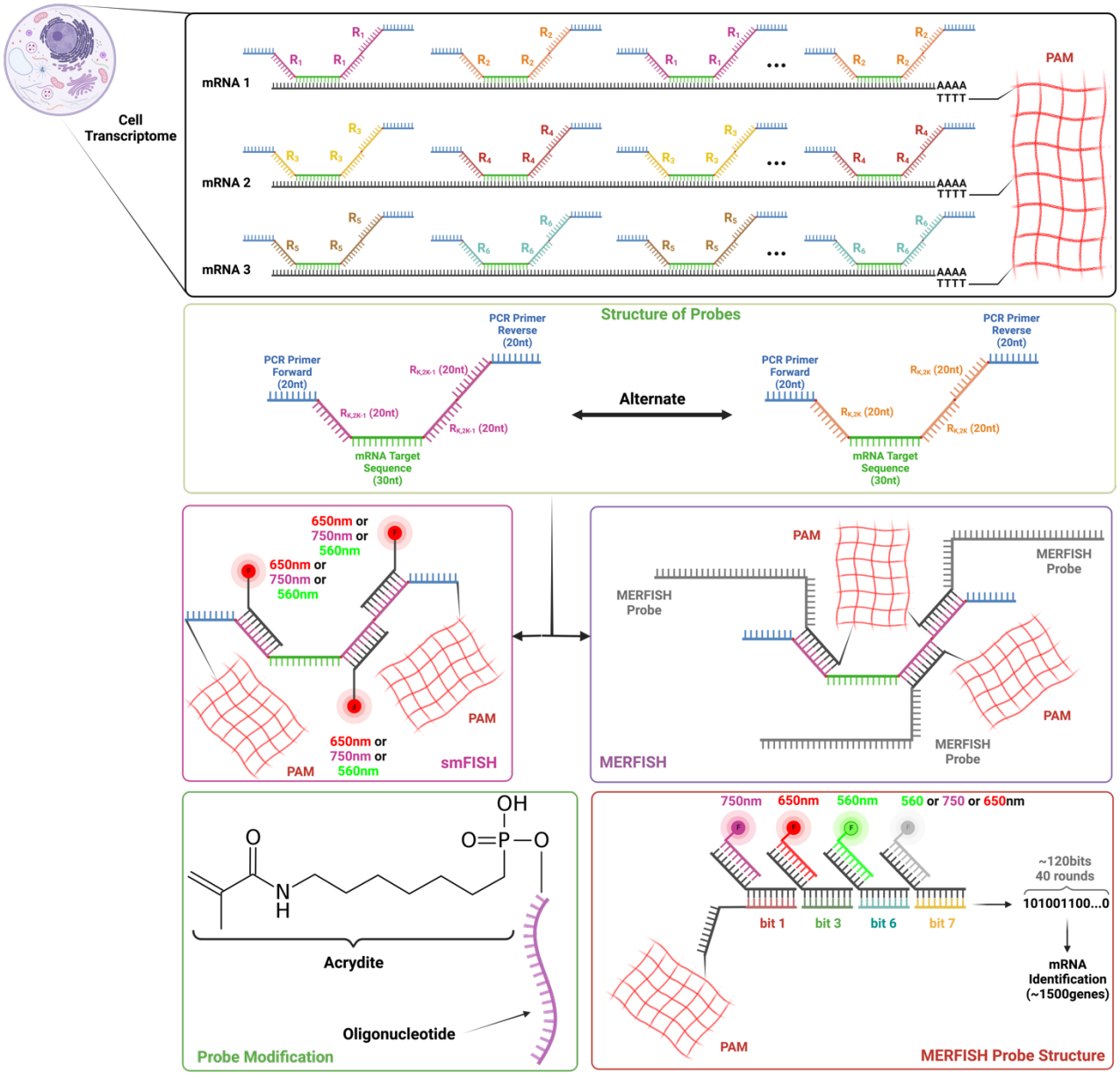

Fig. 2.10: Encoding strategy and probe design.

As shown in Fig. 2.10, we have designed 60 primary probes (30 per readout sequence type) for each mRNA target (target genes from our previous analysis). For each mRNA species, two distinct readout sequences have been used to increase the accuracy of RNA molecule efficiency by co-localizing signals, which will be discussed later. mRNA molecules were tethered to the polyacrylamide gel using their polyA tails to minimize their drift during the imaging process. Each primary probe has a 30nt length target sequence designed complementary to the RNA strand. Moreover, each primary probe has three readout sequences which we used later for mRNA detection using fluorophore-conjugated oligos. To amplify our library, we also put PCR primers at each end of all primary probes, which enabled us to amplify primary probes to maintain our storage. We used two distinct readout sequences per mRNA species and tried to alternate between these two while hybridizing the primary probes to the mRNA molecule. Also, to stabilize and tether the primary probes to the matrix, we used acrydite-conjugated oligos. Due to their molecular structure, acrydite-conjugated oligos could be incorporated into the hydrogel (polyacrylamide), resulting in lower drift between frames and more accuracy in mRNA detection.

From a chemical perspective, the double-bond in the molecular structure of the acrydite group is similar to the double bond of acrylamide monomers and reacts with activated double bonds of acrylamide and bisacrylamide (for cross-linking linear polymers of acrylamide) monomers which results in the incorporation and tethering of acrydite-conjugated oligos in the matrix. By hybridizing the primary probes, both smFISH and MERFISH could be performed to determine the transcriptome of the cells and classify the cell type. Although MERFISH has the advantage of being multiplexed, which means lower imaging time and data storage size, the detection accuracy of smFISH is superior to that of MERFISH. As a result, we decided to choose 289 target genes and perform both smFISH and MERFISH as a comparison. Moreover, we planned

to use the smFISH as ground truth to design the currently developing MERFISH encoding strategy. We chose 289 genes in a way to ensure that we would be able to detect significant cell types across the mice's brains.

For the smFISH experiment, since there are two distinct readout sequences per mRNA species, we could use co-localization of signals to increase the detection accuracy. As shown in Fig. 2.10, the fluorophore-conjugated oligos in three colors (750nm, 650nm, 560nm) were used for imaging purposes.

The currently developing secondary MERFISH probes could be hybridized with the primary ones. Each MERFISH probe is designed in a way that it has four on-bit regions (20nt each), which could be used for detection as shown in Fig. 2.10. However, we will concentrate on the smFISH results in the next chapter.

2.4.Methods

We have sacrificed two 8-week-old C57BL/6J male mice without brain perfusion. Next, we covered the brains with NEG-50 frozen section medium and put them on dry ice for 30mins until they were fully frozen. Next, we sectioned the brain at 16°C and 16μm thickness using a cryo-stat machine and obtained two half-brain sections on treated coverslips. To prepare the coverslips, we mixed 350mL of 37% HCl and 350mL of 100% methanol in a large beaker. Then we put the coverslips in the large beaker by loading them into wafer boats. Coverslips were incubated in the HCl/methanol solution for 30mins. Next, we washed the coverslips three times with ddH_2O by dipping the wafer boats in deionized water. After the third wash, any remaining water wiped out very carefully. Then, we put the wafer boats containing the coverslips into 100% ethanol for 30mins. Next, we took out the coverslips and incubate them at 37°C overnight.

To add an acrylamide-reactive silane layer to the surface of coverslips to allow a polyacrylamide gel to chemically bond to the surfaces of the coverslips, we silanized the coverslips the next day. To do so, we added 700mL of chloroform to 0.7mL of Triethylamine in a large glass beaker and mixed thoroughly. Then we added 1.4mL of Allytrichlorosilane to the solution and mixed vigorously. Then, we put the coverslips into the wafer boats and submerged them in the solution for 30mins at room temperature. Then, in another large glassware, we added 700mL of 100% chloroform and dipped the coverslips and wafer boats into it for a brief wash. Then, in another glassware, we put 700mL of 100% ethanol and briefly washed the coverslips by dipping the boat inside the ethanol. Then we air-dried the coverslips and incubated them at 60°C for 2 hours before coating them with Poly-L-Lysine.

Surfaces of the coverslips become hydrophobic due to Allyltrichlorosilane treatment. As a result, we treated the coverslips with poly-L-lysine and RNase inhibitors to make them hydrophilic so that the tissue could attach to the coverslip. To do so, we added 10mL of poly-L-lysine with 2μL of RNase inhibitor into a 15 mL tube and mixed it thoroughly. Then we put each coverslip into a dish and added the mixture to cover the surface of the coverslips and incubated at 60°C for 30mins. After this step, coverslips are ready for brain sectioning.

For the probe hybridization step, we used a similar protocol as discussed in [Chen 2015].

CHAPTER 3: RESULTS AND DISCUSSIONS

3.1.smFISH Results

We performed smFISH using the designed probes discussed in Chapter 2 and a home-built microscope on an 8-week-old C57BL/6J male mice brain. After sacrificing the mice, we sectioned the brain and produced two slices, 14μm in thickness each, and put them on the poly-lysine-treated coverslips. To image the whole mouse brain slices, we divided the brain slices into distinct fields of view (FoVs), as shown in Fig. 3.1.

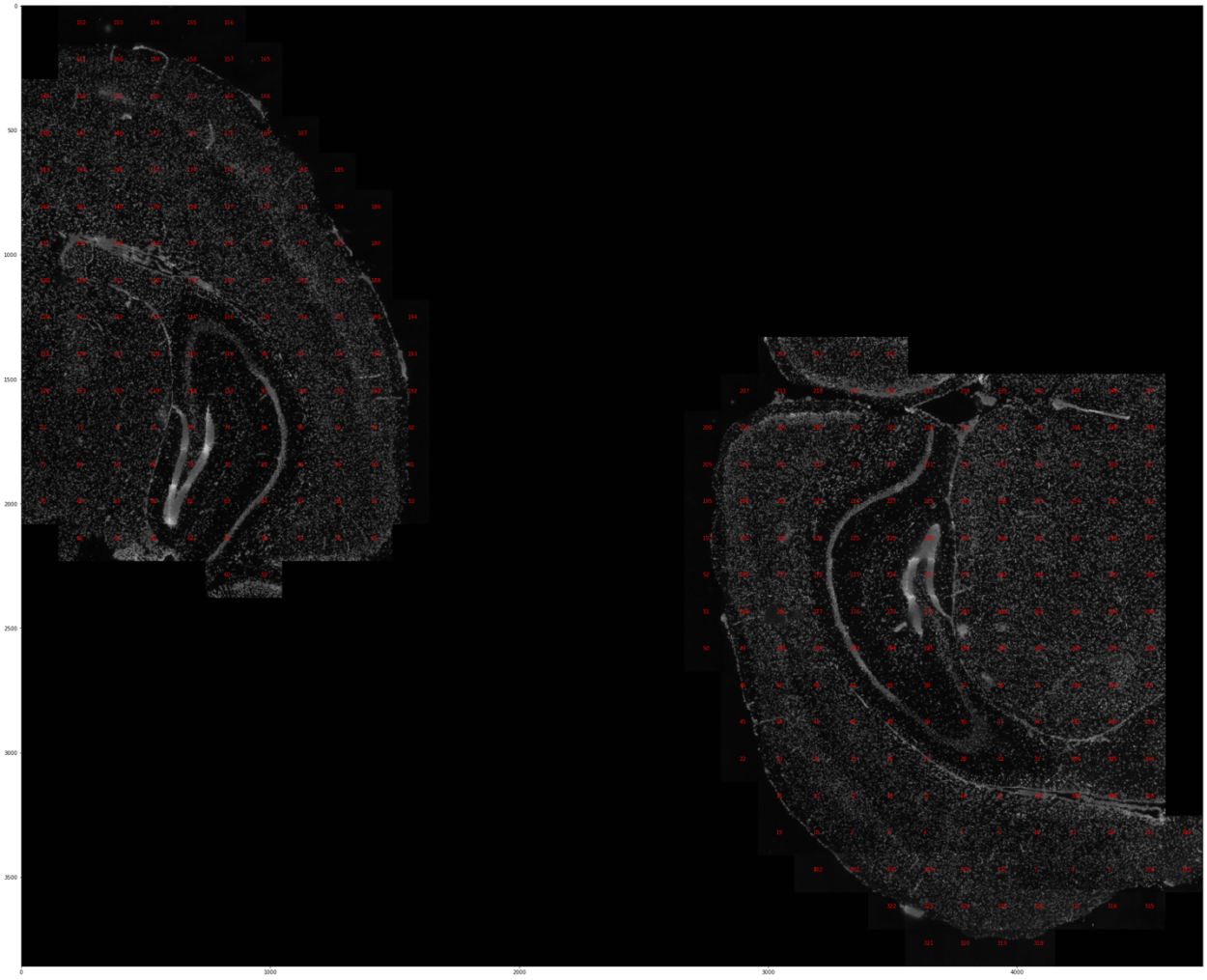

Fig. 3.1: FoVs orientations.

Later, by putting all FoVs together, we could be able to construct the transcriptome of cells in the brain slice. However, here we just show FoV=79 which shows part of the granular layer of the dentate gyrus.

We have designed primary probes with two readout sequences per each type of mRNA molecule. As shown in Fig. 3.2 for *Ascl1*, it is expected to see co-localization of the two signals after imaging each readout sequence. As a result, if two spots co-localize, the certainty of calling that mRNA molecule would be higher. Simply put, by using 2-color smFISH, we could be able to decrease the false positives.

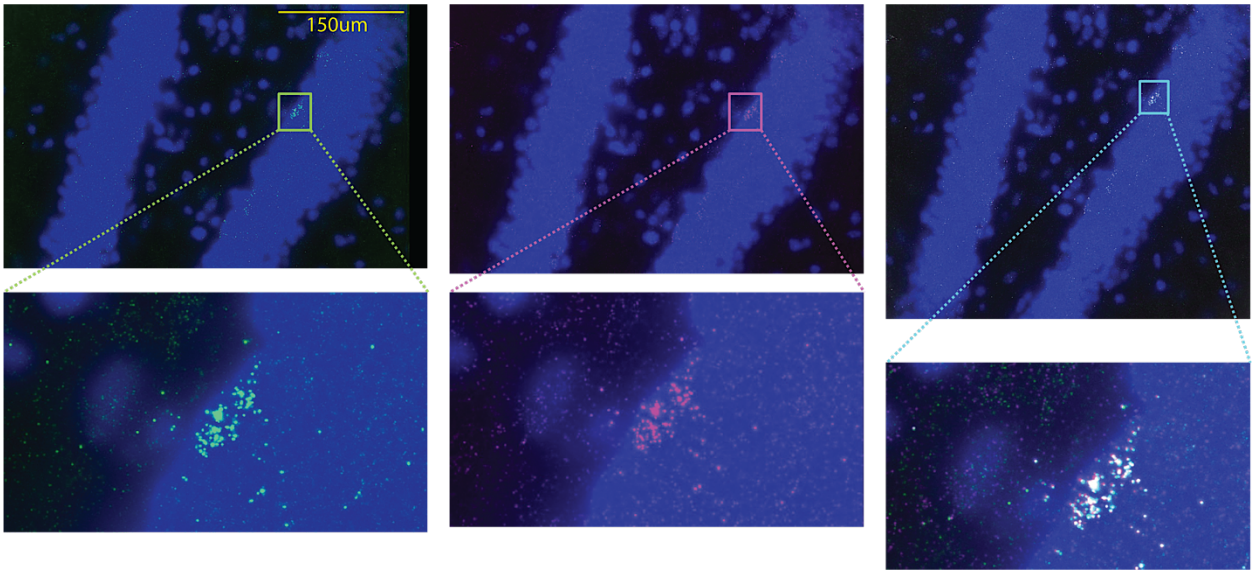

Fig. 3.2: 2-color smFISH for *Ascl1*.

Fig. 3.3 shows some of the highest expressed genes in the mice brain, such as *Malat1* and *Aldoc*. As expected, these genes are expressed everywhere and in all cells. This expression pattern is consistent with previously reported expression from single-cell/nucleus RNA-seq datasets. Moreover, expression of *Neurod2,* which is a transcription factor playing a role in the development of the central and peripheral nervous systems, could be observed in cortical neurons and the

granular layer of the dentate gyrus. Also, as expected, expression of the *Htt* gene was observed everywhere.

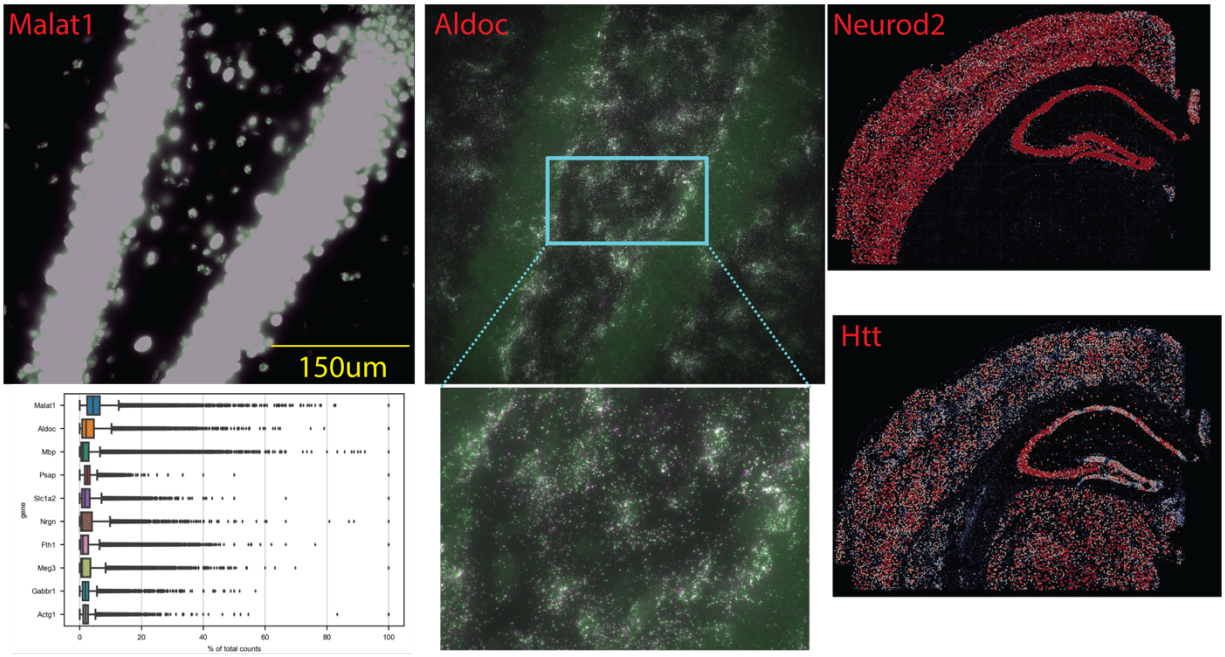

Fig. 3.3: Highly expressed genes.

All these results are in perfect agreement with biological knowledge, which shows the accuracy and sensitivity of our approach. Fig. 3.4 shows the expression of *Dcx,* a marker gene for immature neurons.

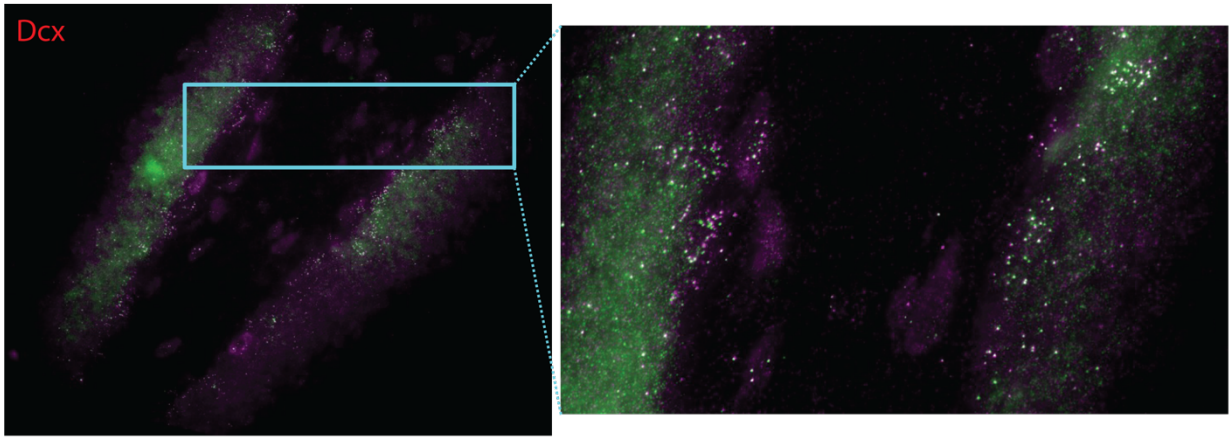

Fig. 3.4: Expression pattern of *Dcx*.

As shown in Fig. 3.4, some cells in the dentate gyrus express *Dcx,* which supports the hypothesis that neurogenesis is still ongoing in young mice. Moreover, many mRNA molecules show the co-localization of both readout sequences, which offers detection accuracy.

Fig. 3.5 shows the expression of *Gfap,* a marker for glial cells and astrocytes. As expected, these glial cells are present in the dentate gyrus.

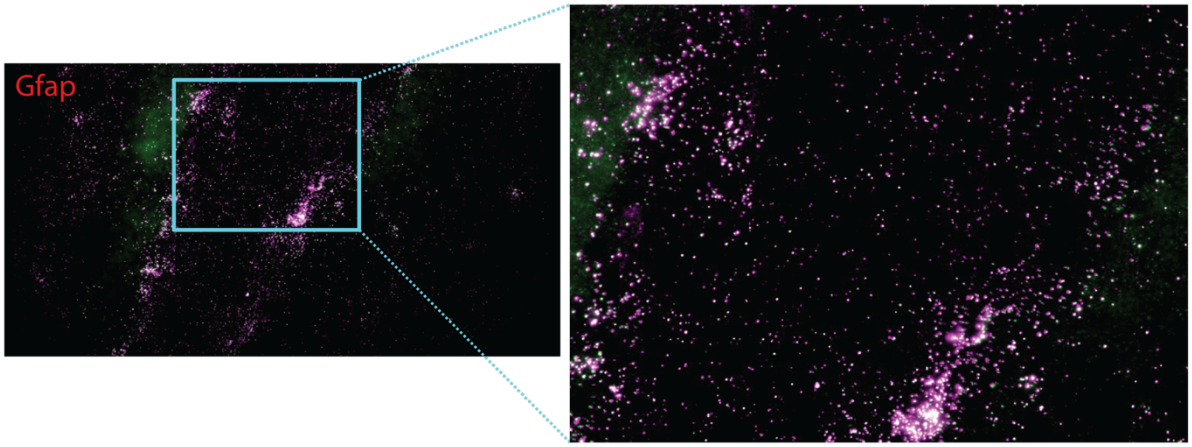

Fig. 3.5: Expression pattern of *Gfap*.

Fig. 3.6 shows the expression of *Olig1,* a marker for oligodendrocytes. Fewer yet localized expressions of this gene are observed in the dentate gyrus.

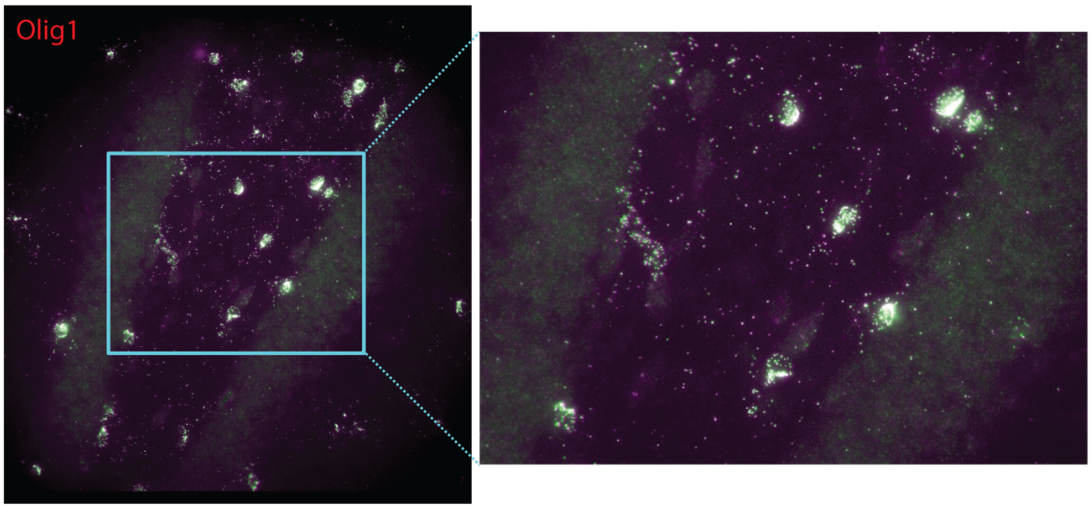

Fig. 3.6: Expression pattern of *Olig1*.

Next, to better understand how good our smFISH experiment was, we compared our results with two previously done MERFISH experiments from our group. In the first MERFISH experiment on the 8-week mouse, 1202transcripts per cell were detected. In the second MERFISH on the 1-year-old mouse, a 1067 transcript was identified. And finally, in our 2-color smFISH, we identified 1035 transcripts. As shown in Fig. 3.7, we also compared transcripts per cell for almost 220 genes and observed a very high correlation (p is a Pearson correlation coefficient). This shows the reproducibility of MERISH and smFISH experiments.

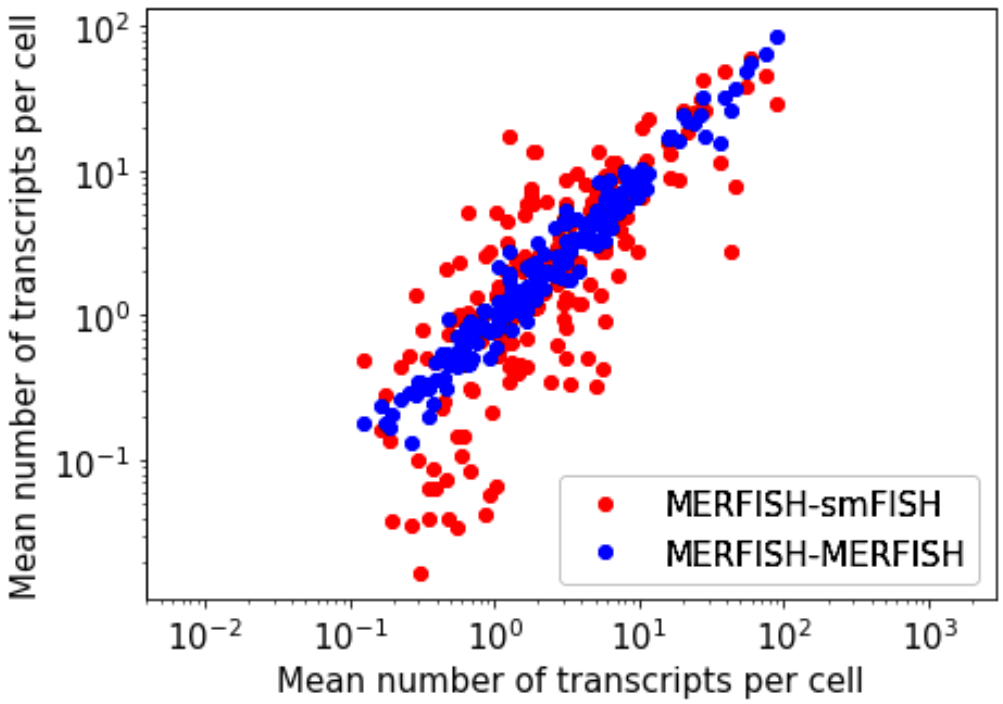

Fig. 3.7: Correlation of various MERFISH and smFISH experiments.

We also compared the correlation between the two-readout sequence we designed, as shown in Fig. 3.8. For instance, in the case of *Prox1* and *Gfap* genes, we saw a very nice correlation which ensures the high mRNA detection efficiency and reliability of our data. This means that we successfully detected each mRNA with both readout sequences.

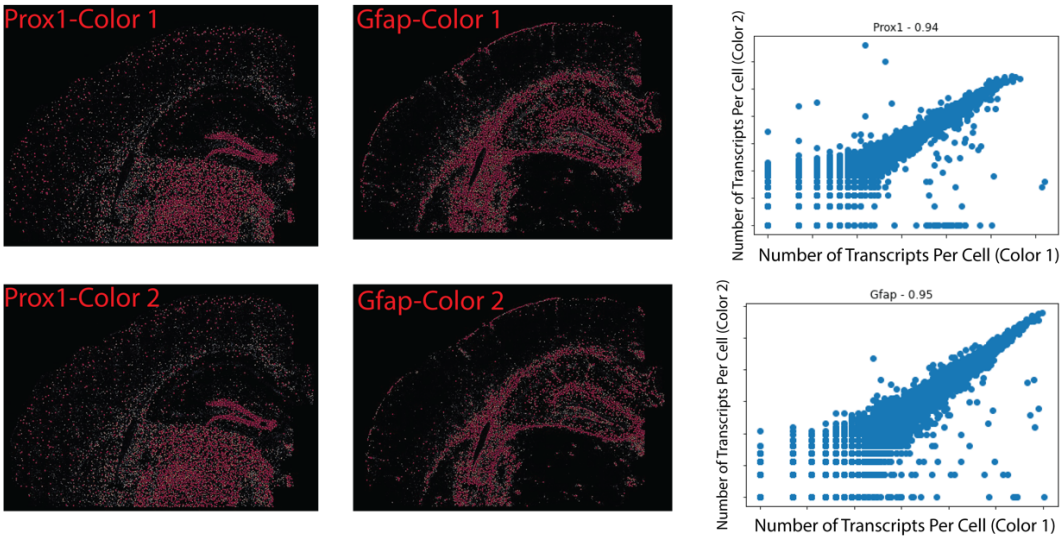

Fig. 3.8: Correlation of 2-color smFISH readouts.

After sequentially imaging the 289 genes, we identified the cell types and clustered them by transferring the labels from [Zeisel 2018] using the Scanpy python package [Wolf 2018]. Fig. 3.9 shows the spatial information of the cell types across the imaged brain slices.

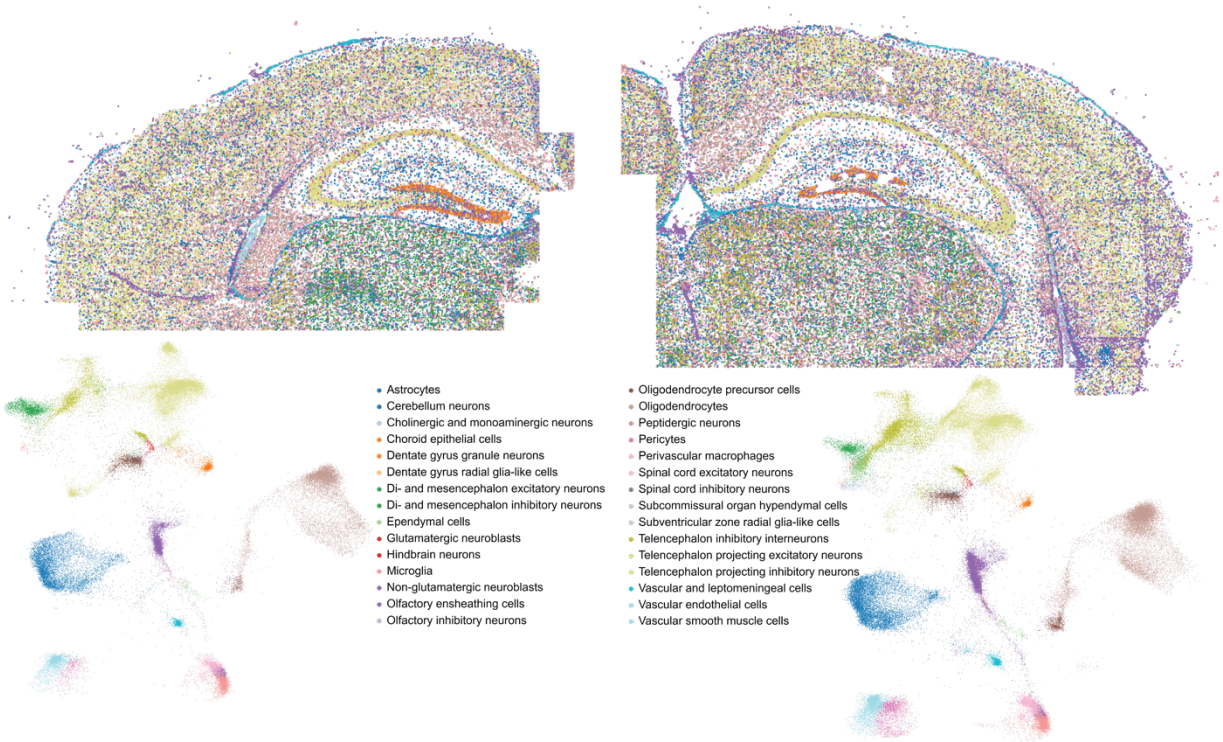

Fig. 3.9: Spatial distribution of the profiled cell types across the brain slices.

As shown in Fig. 3.8, the spatial information of the cell types shows agreement with biology knowledge. For instance, as expected, the granular layer of the dentate gyrus is mainly composed of neurons and astrocytes. However, more genes are needed to be imaged to fully distinguish the cortex layers or subtypes of the major cell types. In Fig. 3.6, the UMAP associated with each brain slice is also shown to prove the superiority of MERFISH to the single-cell/nucleus RNA-seq approach in which spatial information is lost.

Although we have imaged incomplete slices of the brain (two pieces), we can continue the same approach to many other brain slices and then connect them to profile the cell types in the mouse brain in a pan-brain fashion. Moreover, by imaging the complete set of 1500genes, we could better understand the brain composition and strive to complete the periodic table of the brain cell types.

REFERENCES

Hooke, R., 1665. Micrographia. The Royal Society.

Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, Chen H, Wang J, Tang H, Ge W, Zhou Y, Ye F, Jiang M, Wu J, Xiao Y, Jia X, Zhang T, Ma X, Zhang Q, Bai X, Lai S, Yu C, Zhu L, Lin R, Gao Y, Wang M, Wu Y, Zhang J, Zhan R, Zhu S, Hu H, Wang C, Chen M, Huang H, Liang T, Chen J, Wang W, Zhang D, Guo G. Construction of a human cell landscape at the single-cell level. Nature. 2020 May;581(7808):303-309. DOI: 10.1038/s41586-020-2157-4. Epub 2020 Mar 25. PMID: 32214235.

Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, Häring M, Braun E, Borm LE, La Manno G, Codeluppi S, Furlan A, Lee K, Skene N, Harris KD, Hjerling-Leffler J, Arenas E, Ernfors P, Marklund U, Linnarsson S. Molecular Architecture of the Mouse Nervous System. *Cell*. 2018 Aug 9;174(4):999-1014.e22. doi: 10.1016/j.cell.2018.06.021. PMID: 30096314; PMCID: PMC6086934.

Fishell G, Heintz N. The neuron identity problem: form meets function. Neuron. 2013 Oct 30;80(3):602-12. DOI: 10.1016/j.neuron.2013.10.035. PMID: 24183013.

Mukamel EA, Ngai J. Perspectives on defining cell types in the brain. Curr Opin Neurobiol. 2019 Jun;56:61-68. DOI: 10.1016/j.conb.2018.11.007. Epub 2018 Dec 6. PMID: 30530112; PMCID: PMC6551297.

Nelson SB, Sugino K, Hempel CM. The problem of neuronal cell types: a physiological genomics approach. Trends Neurosci. 2006 Jun;29(6):339-45. DOI: 10.1016/j.tins.2006.05.004. Epub 2006 May 22. PMID: 16714064.

Seung HS, Sümbül U. Neuronal cell types and connectivity: lessons from the retina. Neuron. 2014 Sep 17;83(6):1262-72. DOI: 10.1016/j.neuron.2014.08.054. PMID: 25233310; PMCID: PMC4206525.

Zeng H, Sanes JR. Neuronal cell-type classification: challenges, opportunities and the path forward. Nat Rev Neurosci. 2017 Sep;18(9):530-546. DOI: 10.1038/nrn.2017.85. Epub 2017 Aug 3. PMID: 28775344.

Zeng H. What is a cell type, and how to define it? Cell. 2022 Jul 21;185(15):2739-2755. DOI: 10.1016/j.cell.2022.06.031. PMID: 35868277; PMCID: PMC9342916.

Mund A, Coscia F, Kriston A, Hollandi R, Kovács F, Brunner AD, Migh E, Schweizer L, Santos A, Bzorek M, Naimy S, Rahbek-Gjerdrum LM, Dyring-Andersen B, Bulkescher J, Lukas C, Eckert MA, Lengyel E, Gnann C, Lundberg E, Horvath P, Mann M. Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat Biotechnol. 2022 Aug;40(8):1231-1240. doi: 10.1038/s41587-022-01302-5. Epub 2022 May 19. PMID: 35590073; PMCID: PMC9371970.

Lindeboom RGH, Regev A, Teichmann SA. Towards a Human Cell Atlas: Taking Notes from the Past. Trends Genet. 2021 Jul;37(7):625-630. DOI: 10.1016/j.tig.2021.03.007. Epub 2021 Apr 17. PMID: 33879355.

Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Göttgens B, Hacohen N, Haniffa M, Hemberg M, Kim S, Klenerman P, Kriegstein A, Lein E, Linnarsson S, Lundberg E, Lundeberg J, Majumder P, Marioni JC, Merad M, Mhlanga M, Nawijn M, Netea M, Nolan G, Pe'er D, Phillipakis A, Ponting CP, Quake S, Reik W, Rozenblatt-Rosen O, Sanes J, Satija R, Schumacher TN, Shalek A, Shapiro E, Sharma P, Shin JW, Stegle O, Stratton M, Stubbington MJT, Theis FJ, Uhlen M, van Oudenaarden A, Wagner A, Watt F, Weissman J, Wold B, Xavier R, Yosef N; Human Cell Atlas Meeting Participants. The Human Cell Atlas. Elife. 2017 Dec 5;6:e27041. doi: 10.7554/eLife.27041. PMID: 29206104; PMCID: PMC5762154.

BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. Nature. 2021 Oct;598(7879):86-102. DOI: 10.1038/s41586-021-03950-0. Epub 2021 Oct 6. PMID: 34616075; PMCID: PMC8494634.

Ngai J. BRAIN 2.0: Transforming neuroscience. Cell. 2022 Jan 6;185(1):4-8. DOI: 10.1016/j.cell.2021.11.037. PMID: 34995517.

Close JL, Long BR, Zeng H. Spatially resolved transcriptomics in neuroscience. Nat Methods. 2021 Jan;18(1):23-25. DOI: 10.1038/s41592-020-01040-z. PMID: 33408398.

Larsson L, Frisén J, Lundeberg J. Spatially resolved transcriptomics adds a new dimension to genomics. Nat Methods. 2021 Jan;18(1):15-18. DOI: 10.1038/s41592-020-01038-7. PMID: 33408402.

Moses L, Pachter L. Museum of spatial transcriptomics. Nat Methods. 2022 May;19(5):534-546. DOI: 10.1038/s41592-022-01409-2. Epub 2022 Mar 10. Erratum in: Nat Methods. 2022 Apr 19;: PMID: 35273392.

Zhuang X. Spatially resolved single-cell genomics and transcriptomics by imaging. Nat Methods. 2021 Jan;18(1):18-22. DOI: 10.1038/s41592-020-01037-8. PMID: 33408406.

Moffitt JR, Bambah-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, Perez JD, Rubinstein ND, Hao J, Regev A, Dulac C, Zhuang X. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science. 2018 Nov 16;362(6416):eaau5324. DOI: 10.1126/science.aau5324. Epub 2018 Nov 1. PMID: 30385464; PMCID: PMC6482113.

Cho NH, Cheveralls KC, Brunner AD, Kim K, Michaelis AC, Raghavan P, Kobayashi H, Savy L, Li JY, Canaj H, Kim JYS, Stewart EM, Gnann C, McCarthy F, Cabrera JP, Brunetti RM, Chhun BB, Dingle G, Hein MY, Huang B, Mehta SB, Weissman JS, Gómez-Sjöberg R, Itzhak DN, Royer LA, Mann M, Leonetti MD. OpenCell: Endogenous tagging for the cartography of

human cellular organization. Science. 2022 Mar 11;375(6585):eabi6983. DOI: 10.1126/science.abi6983. Epub 2022 Mar 11. PMID: 35271311; PMCID: PMC9119736.

Gao L, Liu S, Gou L, Hu Y, Liu Y, Deng L, Ma D, Wang H, Yang Q, Chen Z, Liu D, Qiu S, Wang X, Wang D, Wang X, Ren B, Liu Q, Chen T, Shi X, Yao H, Xu C, Li CT, Sun Y, Li A, Luo Q, Gong H, Xu N, Yan J. Single-neuron projection of mouse prefrontal cortex. Nat Neurosci. 2022 Apr;25(4):515-529. DOI: 10.1038/s41593-022-01041-5. Epub 2022 Mar 31. PMID: 35361973.

Hulse BK, Haberkern H, Franconville R, Turner-Evans DB, Takemura SY, Wolff T, Noorman M, Dreher M, Dan C, Parekh R, Hermundstad AM, Rubin GM, Jayaraman V. A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. Elife. 2021 Oct 26;10:e66039. DOI: 10.7554/eLife.66039. Epub ahead of print. PMID: 34696823.

Südhof TC. The cell biology of synapse formation. J Cell Biol. 2021 Jul 5;220(7):e202103052. DOI: 10.1083/jcb.202103052. Epub 2021 Jun 4. PMID: 34086051; PMCID: PMC8186004.

Trotter JH, Hao J, Maxeiner S, Tsetsenis T, Liu Z, Zhuang X, Südhof TC. Synaptic neurexin-1 assembles into dynamically regulated active zone nanoclusters. J Cell Biol. 2019 Aug 5;218(8):2677-2698. DOI: 10.1083/jcb.201812076. Epub 2019 Jul 1. PMID: 31262725; PMCID: PMC6683742.

Zhu C, Preissl S, Ren B. Single-cell multimodal omics: the power of many. Nat Methods. 2020 Jan;17(1):11-14. DOI: 10.1038/s41592-019-0691-5. PMID: 31907462.

Berg J, Sorensen SA, Ting JT, Miller JA, Chartrand T, Buchin A, Bakken TE, Budzillo A, Dee N, Ding SL, Gouwens NW, Hodge RD, Kalmbach B, Lee C, Lee BR, Alfiler L, Baker K, Barkan E, Beller A, Berry K, Bertagnolli D, Bickley K, Bomben J, Braun T, Brouner K, Casper T, Chong P, Crichton K, Dalley R, de Frates R, Desta T, Lee SD, D'Orazi F, Dotson N, Egdorf T, Enstrom R, Farrell C, Feng D, Fong O, Furdan S, Galakhova AA, Gamlin C, Gary A, Glandon A, Goldy J, Gorham M, Goriounova NA, Gratiy S, Graybuck L, Gu H, Hadley K, Hansen N, Heistek TS, Henry AM, Heyer DB, Hill D, Hill C, Hupp M, Jarsky T, Kebede S, Keene L, Kim L, Kim MH, Kroll M, Latimer C, Levi BP, Link KE, Mallory M, Mann R, Marshall D, Maxwell M, McGraw M, McMillen D, Melief E, Mertens EJ, Mezei L, Mihut N, Mok S, Molnar G, Mukora A, Ng L, Ngo K, Nicovich PR, Nyhus J, Olah G, Oldre A, Omstead V, Ozsvar A, Park D, Peng H, Pham T, Pom CA, Potekhina L, Rajanbabu R, Ransford S, Reid D, Rimorin C, Ruiz A, Sandman D, Sulc J, Sunkin SM, Szafer A, Szemenyei V, Thomsen ER, Tieu M, Torkelson A, Trinh J, Tung H, Wakeman W, Waleboer F, Ward K, Wilbers R, Williams G, Yao Z, Yoon JG, Anastassiou C, Arkhipov A, Barzo P, Bernard A, Cobbs C, de Witt Hamer PC, Ellenbogen RG, Esposito L, Ferreira M, Gwinn RP, Hawrylycz MJ, Hof PR, Idema S, Jones AR, Keene CD, Ko AL, Murphy GJ, Ng L, Ojemann JG, Patel AP, Phillips JW, Silbergeld DL, Smith K, Tasic B, Yuste R, Segev I, de Kock CPJ, Mansvelder HD, Tamas G, Zeng H, Koch C, Lein ES. The human neocortical expansion involves glutamatergic neuron diversification. Nature. 2021 Oct;598(7879):151-158. DOI: 10.1038/s41586-021-03813-8. Epub 2021 Oct 6. Erratum in: Nature. 2022 Jan;601(7893): E12. PMID: 34616067; PMCID: PMC8494638.

Scala F, Kobak D, Bernabucci M, Bernaerts Y, Cadwell CR, Castro JR, Hartmanis L, Jiang X, Laturnus S, Miranda E, Mulherkar S, Tan ZH, Yao Z, Zeng H, Sandberg R, Berens P, Tolias AS. Phenotypic variation of transcriptomic cell types in mouse motor cortex. Nature. 2021 Oct;598(7879):144-150. DOI: 10.1038/s41586-020-2907-3. Epub 2020 Nov 12. PMID: 33184512; PMCID: PMC8113357.

Le Floch P, Li Q, Lin Z, Zhao S, Liu R, Tasnim K, Jiang H, Liu J. Stretchable Mesh Nanoelectronics for 3D Single-Cell Chronic Electrophysiology from Developing Brain Organoids. Adv Mater. 2022 Mar;34(11):e2106829. DOI: 10.1002/adma.202106829. Epub 2022 Feb 6. PMID: 35014735; PMCID: PMC8930507.

Kim EJ, Zhang Z, Huang L, Ito-Cole T, Jacobs MW, Juavinett AL, Senturk G, Hu M, Ku M, Ecker JR, Callaway EM. Extraction of Distinct Neuronal Cell Types from within a Genetically Continuous Population. Neuron. 2020 Jul 22;107(2):274-282.e6. DOI: 10.1016/j.neuron.2020.04.018. Epub 2020 May 11. PMID: 32396852; PMCID: PMC7381365.

Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller J, Fong O, Hirokawa KE, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova NV, Hirschstein D, Pendergraft J, Sullivan HA, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H. Shared and distinct transcriptomic cell types across neocortical areas. Nature. 2018 Nov;563(7729):72-78. DOI: 10.1038/s41586-018-0654-5. Epub 2018 Oct 31. PMID: 30382198; PMCID: PMC6456269.

Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, Dong H, Zhuang X. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. Nature. 2021 Oct;598(7879):137-143. DOI: 10.1038/s41586-021-03705-x. Epub 2021 Oct 6. PMID: 34616063; PMCID: PMC8494645.

Burgeon S, Duffield J, Dipoppa M, Ritoux A, Prankerd I, Nicoloutsopoulos D, Orme D, Shinn M, Peng H, Forrest H, Viduolyte A, Reddy CB, Isogai Y, Carandini M, Harris KD. A transcriptomic axis predicts state modulation of cortical interneurons. Nature. 2022 Jul;607(7918):330-338. DOI: 10.1038/s41586-022-04915-7. Epub 2022 Jul 6. Erratum in: Nature. 2022 Aug 25;: PMID: 35794483; PMCID: PMC9279161.

Condyles C, Ghanbari A, Manjrekar N, Bistrong K, Yao S, Yao Z, Nguyen TN, Zeng H, Tasic B, Chen JL. Dense functional and molecular readout of a circuit hub in the sensory cortex. Science. 2022 Jan 7;375(6576):eabl5981. DOI: 10.1126/science.abl5981. Epub 2022 Jan 7. PMID: 34990233; PMCID: PMC9255671.

Lovett-Barron M, Chen R, Bradbury S, Andalman AS, Wagle M, Guo S, Deisseroth K. Multiple convergent hypothalamus-brainstem circuits drive defensive behavior. Nat Neurosci. 2020

Aug;23(8):959-967. DOI: 10.1038/s41593-020-0655-1. Epub 2020 Jun 22. PMID: 32572237; PMCID: PMC7687349.

Xu S, Yang H, Menon V, Lemire AL, Wang L, Henry FE, Turaga SC, Sternson SM. Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. Science. 2020 Oct 16;370(6514):eabb2494. DOI: 10.1126/science.abb2494. PMID: 33060330.

Yao Z, Liu H, Xie F, Fischer S, Adkins RS, Aldridge AI, Ament SA, Bartlett A, Behrens MM, Van den Berge K, Bertagnolli D, de Bézieux HR, Biancalani T, Booeshaghi AS, Bravo HC, Casper T, Colantuoni C, Crabtree J, Creasy H, Crichton K, Crow M, Dee N, Dougherty EL, Doyle WI, Dudoit S, Fang R, Felix V, Fong O, Giglio M, Goldy J, Hawrylycz M, Herb BR, Hertzano R, Hou X, Hu Q, Kancherla J, Kroll M, Lathia K, Li YE, Lucero JD, Luo C, Mahurkar A, McMillen D, Nadaf NM, Nery JR, Nguyen TN, Niu SY, Ntranos V, Orvis J, Osteen JK, Pham T, Pinto-Duarte A, Poirion O, Preissl S, Purdom E, Rimorin C, Risso D, Rivkin AC, Smith K, Street K, Sulc J, Svensson V, Tieu M, Torkelson A, Tung H, Vaishnav ED, Vanderburg CR, van Velthoven C, Wang X, White OR, Huang ZJ, Kharchenko PV, Pachter L, Ngai J, Regev A, Tasic B, Welch JD, Gillis J, Macosko EZ, Ren B, Ecker JR, Zeng H, Mukamel EA. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. Nature. 2021 Oct;598(7879):103-110. DOI: 10.1038/s41586-021-03500-8. Epub 2021 Oct 6. PMID: 34616066; PMCID: PMC8494649.

Daigle TL, Madisen L, Hage TA, Valley MT, Knoblich U, Larsen RS, Takeno MM, Huang L, Gu H, Larsen R, Mills M, Bosma-Moody A, Siverts LA, Walker M, Graybuck LT, Yao Z, Fong O, Nguyen TN, Garren E, Lenz GH, Chavarha M, Pendergraft J, Harrington J, Hirokawa KE, Harris JA, Nicovich PR, McGraw MJ, Ollerenshaw DR, Smith KA, Baker CA, Ting JT, Sunkin SM, Lecoq J, Lin MZ, Boyden ES, Murphy GJ, da Costa NM, Waters J, Li L, Tasic B, Zeng H. A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. Cell. 2018 Jul 12;174(2):465-480.e22. DOI: 10.1016/j.cell.2018.06.035. PMID: 30007418; PMCID: PMC6086366.

Matho KS, Huilgol D, Galbavy W, He M, Kim G, An X, Lu J, Wu P, Di Bella DJ, Shetty AS, Palaniswamy R, Hatfield J, Raudales R, Narasimhan A, Gamache E, Levine JM, Tucciarone J, Szelenyi E, Harris JA, Mitra PP, Osten P, Arlotta P, Huang ZJ. Genetic dissection of the glutamatergic neuron system in the cerebral cortex. Nature. 2021 Oct;598(7879):182-187. DOI: 10.1038/s41586-021-03955-9. Epub 2021 Oct 6. PMID: 34616069; PMCID: PMC8494647.

Morris SA. Cell identity reprogrammed. Nature. 2019 Nov;575(7781):44-45. DOI: 10.1038/d41586-019-02834-8. PMID: 31686043.

Briggs R, King TJ. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. Proc Natl Acad Sci U A. 1952 May;38(5):455-63. DOI: 10.1073/pnas.38.5.455. PMID: 16589125; PMCID: PMC1063586.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006 Aug 25;126(4):663-76. DOI: 10.1016/j.cell.2006.07.024. Epub 2006 Aug 10. PMID: 16904174.

Gurdon J. B., Elsdale T. R., Fischberg M. Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature. 1958 Jul 5;182(4627):64-5. DOI: 10.1038/182064a0. PMID: 13566187.

Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015 Apr 24;348(6233):aaa6090. DOI: 10.1126/science.aaa6090. Epub 2015 Apr 9. PMID: 25858977; PMCID: PMC4662681.

Yao Z, Liu H, Xie F, Fischer S, Adkins RS, Aldridge AI, Ament SA, Bartlett A, Behrens MM, Van den Berge K, Bertagnolli D, de Bézieux HR, Biancalani T, Booeshaghi AS, Bravo HC, Casper T, Colantuoni C, Crabtree J, Creasy H, Crichton K, Crow M, Dee N, Dougherty EL, Doyle WI, Dudoit S, Fang R, Felix V, Fong O, Giglio M, Goldy J, Hawrylycz M, Herb BR, Hertzano R, Hou X, Hu Q, Kancherla J, Kroll M, Lathia K, Li YE, Lucero JD, Luo C, Mahurkar A, McMillen D, Nadaf NM, Nery JR, Nguyen TN, Niu SY, Ntranos V, Orvis J, Osteen JK, Pham T, Pinto-Duarte A, Poirion O, Preissl S, Purdom E, Rimorin C, Risso D, Rivkin AC, Smith K, Street K, Sulc J, Svensson V, Tieu M, Torkelson A, Tung H, Vaishnav ED, Vanderburg CR, van Velthoven C, Wang X, White OR, Huang ZJ, Kharchenko PV, Pachter L, Ngai J, Regev A, Tasic B, Welch JD, Gillis J, Macosko EZ, Ren B, Ecker JR, Zeng H, Mukamel EA. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. Nature. 2021 Oct;598(7879):103-110. DOI: 10.1038/s41586-021-03500-8. Epub 2021 Oct 6. PMID: 34616066; PMCID: PMC8494649.

Yao Z, van Velthoven CTJ, Nguyen TN, Goldy J, Sedeno-Cortes AE, Baftizadeh F, Bertagnolli D, Casper T, Chiang M, Crichton K, Ding SL, Fong O, Garren E, Glandon A, Gouwens NW, Gray J, Graybuck LT, Hawrylycz MJ, Hirschstein D, Kroll M, Lathia K, Lee C, Levi B, McMillen D, Mok S, Pham T, Ren Q, Rimorin C, Shapovalova N, Sulc J, Sunkin SM, Tieu M, Torkelson A, Tung H, Ward K, Dee N, Smith KA, Tasic B, Zeng H. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell. 2021 Jun 10;184(12):3222-3241.e26. DOI: 10.1016/j.cell.2021.04.021. Epub 2021 May 17. PMID: 34004146; PMCID: PMC8195859.

Di Bella DJ, Habibi E, Stickels RR, Scalia G, Brown J, Yadollahpour P, Yang SM, Abbate C, Biancalani T, Macosko EZ, Chen F, Regev A, Arlotta P. Molecular logic of cellular diversification in the mouse cerebral cortex. Nature. 2021 Jul;595(7868):554-559. DOI: 10.1038/s41586-021-03670-5. Epub 2021 Jun 23. Erratum in: Nature. 2021 Aug;596(7873): E11. PMID: 34163074; PMCID: PMC9006333.

Hochgerner H, Zeisel A, Lönnerberg P, Linnarsson S. Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. Nat Neurosci. 2018 Feb;21(2):290-299. DOI: 10.1038/s41593-017-0056-2. Epub 2018 Jan 15. PMID: 29335606.

Saunders A, Macosko EZ, Wysoker A, Goldman M, Krienen FM, de Rivera H, Bien E, Baum M, Bortolin L, Wang S, Goeva A, Nemesh J, Kamitaki N, Brumbaugh S, Kulp D, McCarroll SA. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell. 2018

Aug 9;174(4):1015-1030.e16. DOI: 10.1016/j.cell.2018.07.028. PMID: 30096299; PMCID: PMC6447408.

La Manno G, Siletti K, Furlan A, Gyllborg D, Vinsland E, Mossi Albiach A, Mattsson Langseth C, Khven I, Lederer AR, Dratva LM, Johnsson A, Nilsson M, Lönnerberg P, Linnarsson S. Molecular architecture of the developing mouse brain. Nature. 2021 Aug;596(7870):92-96. doi: 10.1038/s41586-021-03775-x. Epub 2021 Jul 28. PMID: 34321664.

Aevermann B, Zhang Y, Novotny M, Keshk M, Bakken T, Miller J, Hodge R, Lelieveldt B, Lein E, Scheuermann RH. A machine learning method for discovering minimum marker gene combinations for cell type identification from single-cell RNA sequencing. Genome Res. 2021 Oct;31(10):1767-1780. DOI: 10.1101/gr.275569.121. Epub 2021 Jun 4. PMID: 34088715; PMCID: PMC8494219.

Liddelow SA, Guttenplan KA, Clarke LE, Bennett FC, Bohlen CJ, Schirmer L, Bennett ML, Münch AE, Chung WS, Peterson TC, Wilton DK, Frouin A, Napier BA, Panicker N, Kumar M, Buckwalter MS, Rowitch DH, Dawson VL, Dawson TM, Stevens B, Barres BA. Neurotoxic reactive astrocytes are induced by activated microglia. Nature. 2017 Jan 26;541(7638):481-487. DOI: 10.1038/nature21029. Epub 2017 Jan 18. PMID: 28099414; PMCID: PMC5404890.

Qian H, Kang X, Hu J, Zhang D, Liang Z, Meng F, Zhang X, Xue Y, Maimon R, Dowdy SF, Devaraj NK, Zhou Z, Mobley WC, Cleveland DW, Fu XD. Reversing a model of Parkinson's disease with in situ converted nigral neurons. Nature. 2020 Jun;582(7813):550-556. DOI: 10.1038/s41586-020-2388-4. Epub 2020 Jun 24. Erratum in: Nature. 2020 Aug;584(7820): E17. PMID: 32581380; PMCID: PMC7521455.

Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;19(1):15. DOI: 10.1186/s13059-017-1382-0. PMID: 29409532; PMCID: PMC5802054.