

Profiling the dynamics of the active transcriptome in the juvenile and adult brain using spatial transcriptomics

By Ruvimbo Mishi (MSHRUV003)



Supervised by: Dr. Dorit Hockman

Department of Health Sciences

University of Cape Town

South Africa

THESIS PRESENTED FOR THE DEGREE OF

Master of Science

APRIL 2023

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

PLAGARISM DECLARATION

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

I, Ruvimbo Mishi, hereby declare that the work on which this dissertation/thesis is based on my original work (except where acknowledgments indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

Ruvimbo Mishi

4 April 2023

ACKNOWLEDGEMENTS

I extend my sincerest gratitude to my thesis supervisor, Dr. Dorit Hockman, for her invaluable guidance and support throughout my research. Her expertise, mentorship, and dedication were instrumental in shaping the outcome of this thesis.

I would like to acknowledge the financial support that enabled me to conduct this research, which was made possible by the Royal Society Flair fellowship awarded to Dr. Dorit Hockman. I am also grateful to my colleagues and friends for their support and encouragement throughout the process, which helped me to stay motivated and focused.

My heartfelt thanks go to my family for their love and unwavering support. Their encouragement and belief in me have been a constant source of strength throughout this challenging journey.

Finally, I would like to express my gratitude to my partner, Max, for his unwavering support and encouragement throughout this journey. His patience, understanding, and love have been invaluable in helping me to navigate the challenges of completing this thesis. I could not have done this without him.

ABSTRACT

The human brain is made up of a collection of distinct cell types that play specialized roles in maintaining proper brain function. The process of maturation of this complex organ can be assessed by examining the dynamics of gene expression within the various cell types. Previous studies have used single-cell/single nucleus RNA-sequencing (sc/snRNA-seq) to provide vital information on gene expression at a cell-specific level. These studies, which focused on either the pre-natal or adult brain, lacked comprehensive information about the spatial dynamics of cell type-specific gene expression over the course of brain maturation. This information is important for understanding the dynamics of gene expression in the context of changing tissue architecture over the course of maturation. This study aims to contribute to our understanding of the maturing human brain by obtaining the spatial gene expression information of the antemortem brain as it matures from postnatal to adult state.

Samples from the human temporal cortex (4-, 15- (x2), and 31-year-old) were obtained during elective surgeries to treat epilepsy. Tissue samples were freshly frozen in OCT. H&E (hematoxylin and eosin) staining was employed for the initial screening of the tissue samples. The 10x Genomics Visium Spatial gene expression system was used to obtain genome-wide spatial gene expression patterns. The Visium FASTQ files were aligned to the human transcriptome using 10X Genomics *SpaceRanger*. To spatially map the brain cell types, the Visium data was integrated with snRNAseq data using the cell2location method. *In situ* hybridization chain reaction was used to validate the identified gene expression patterns.

Fifty-four cell types, annotated using the current published human temporal lobe cell atlas, were spatially mapped by integrating Visium spatial gene expression and existing snRNA-seq datasets from the same samples. Moreover, this integrative analysis revealed potential changes in the distribution of cell types during brain maturation, highlighting the importance of studying the spatiotemporal dynamics of cell types to better understand brain development and function. Previously identified layer-enriched genes were confirmed to be present in all the samples. Furthermore, layer-enriched genes that showed an increase in expression during brain maturation were identified.

The findings of this study contribute to the human brain cell atlas through the provision of spatial gene expression information in the maturing temporal cortex.

CONTENTS:

CHAPTER 1: LITERATURE REVIEW.....	1
1.1 Introduction	1
1.2 The human brain.....	3
1.3 Different stages of brain development.....	7
1.4 Brain development is under genetic control.	10
1.5 Previous technologies used to obtain gene expression data.	10
1.6 Using single cell and single nucleus RNA sequencing to obtain cell-type specific gene expression data.....	13
1.7 Spatial transcriptomics.....	18
1.8 Validating selected Visium gene expression findings using <i>in situ</i> hybridization chain reaction.	32
1.10 Study Aim	34
1.11 Objectives.....	34
CHAPTER 2: METHODOLOGY.....	35
2.1 Ethical Approval	35
2.2 Sample preparation	35
2.3 Spatial Gene Expression Visium Experiment	36
2.4. The Visium library preparation	45
2.5 Data Analysis.....	52
2.6 In situ Hybridization Chain Reaction (HCR).....	57
CHAPTER 3: RESULTS	60
3.1 Visium Tissue Optimization.....	60
3.2 Spatial transcriptomics using 10X Genomics Visium Spatial Gene expression.....	62
3.3 Library sequencing and alignment.....	63
3.4 Estimating the number of nuclei per spot	65
3.5 Spatial mapping and abundance of 54 cell types across the Visium samples	66
3.6 Identification and verification of layer-specific cell types	71
3.7 Identifying cellular compartments using non-negative matrix factorisation	73
3.8 Leiden clustering.....	77
3.9 Optimization of in situ Hybridization chain reaction (HCR) using positive control probes	84
3.10 Validating the Visium analysis using Bayesspace and HCR.....	88
CHAPTER 4: DISCUSSION	97
4.1 Tissue preservation	97
4.2 Quality control of Visium libraries	97
4.3 Spatially mapping cell types using cell2location.....	98
4.4 Understanding the tissue organization using cell clustering.	100

4.5 Validation of previously identified layer 1 neuronal marker genes, <i>CXCL14</i> and <i>RELN</i> , using Bayesspace and HCR	105
4.6 LIMITATIONS OF STUDY	107
4.7 CONCLUSION.....	108
REFERENCES.....	109

Scripts used in this thesis are found here: <https://github.com/rue1996/Thesis-scripts>

LIST OF FIGURES:

Figure 1. 1 The organization of the brain	4
Figure 1. 2: The basic structure of a neuron.	5
Figure 1. 3: Three different classes of glial cells.	7
Figure 1. 4: Different stages of brain maturation.	8
Figure 1. 5: Schematic showing neurogenesis in the mammalian cortex.	9
Figure 1. 6: The process of bulk RNA sequencing.	12
Figure 1. 7: The process of single-cell RNA sequencing.	15
Figure 1. 8: Laminar organization of the cortex.	19
Figure 1. 9: Spatially resolved transcriptomic technologies.	21
Figure 1. 10: An overview of single-molecule fluorescence in situ hybridization (smFISH) based technologies.	24
Figure 1. 11: An overview of the glass slide used in ST technology where the tissue of interest binds.	26
Figure 1. 12: Illustration of capturing RNA from a tissue using slide-seq	27
Figure 1. 13: Visium Spatial Gene Expression slide.	28
Figure 1. 14: In situ hybridization chain reaction process.	33
 Figure 2. 1: Schematic showing the process of sample preparation.	 36
Figure 2. 2: Schematic representation of the steps in involved in tissue sectioning.	38
Figure 2. 3: Illustration of the H & E staining procedure used.	40
Figure 2. 4: Schematic showing the steps involved in the Visium tissue optimization step.	41
Figure 2. 5: An image of the 10X Genomics slide cassette used in the Visium experiment.	42
Figure 2. 6: An image showing the Visium test imaging slide.	44
Figure 2. 7: The diagram showing the Visium protocol steps & timing.	45
Figure 2. 8: The diagram shows an overview of the data analysis steps used in this study.	52

Figure 3. 1: Visium Tissue Optimization and sectioning.....	61
Figure 3.2: Quality control during the Visium process.....	63
Figure 3. 3: Segmentation to estimate the number of nuclei per spot in the Visium samples.	66
Figure 3.4: Spatial mapping of cell types in the human temporal cortex.	68
Figure 3. 5: Spatial mapping of cell types in the human temporal cortex.	70
Figure 3.6: Spatial mapping of layer-specific cell types.	72
Figure 3. 7: Mapping the cellular compartments using non-negative matrix factorisation (NMF).	75
Figure 3. 8: Mapping the cellular compartments using non-negative matrix factorisation (NMF).	76
Figure 3.9: Identified tissue regions that differ in their cell composition.	81
Figure 3.10: Heatmaps showing the top 20 differentially expressed genes associated with Leiden clusters of interest.....	83
Figure 3. 11: Assessing the effect of hairpin fluorophore on the <i>SLC17A7</i> probe signal on 46- year-old tissue.....	85
Figure 3.12: Assessing the effect of probe concentration on <i>Gad1</i> and <i>Slc17a7</i> probe signal on the 46-year-old.	85
Figure 3.13: Multiplex in situ HCR using positive control probes on the human temporal cortex.	87
Figure 3.14: HCR Tile scan of the temporal cortex.	88
Figure 3.15: Bayesspace and multiplex <i>in situ</i> HCR analysis of layer 1-enriched gene, <i>CXCL14</i> , in temporal cortex tissue sections.	91
Figure 3.16: Bayesspace and multiplex <i>in situ</i> HCR analysis of layer 1-enriched genes <i>RELN</i> , <i>FABP7</i> and <i>AQP4</i> , in temporal cortex tissue sections.....	93
Figure 3. 17: Bayesspace and multiplex in situ HCR analysis of layer 4-6-enriched genes, <i>RORB</i> , <i>CLSTN2</i> and <i>TSHZ2</i> , in temporal cortex tissue sections.....	96

Supplementary Figures 127

Supplementary Figure 1: Tapestation trace indicates the cDNA fragments size distribution after cDNA amplification for the 1st and 2nd Visium experiments.

Supplementary Figure 2: Tapestation trace which shows the fragment size distribution after PCR Indexing for the 1st and 2nd Visium experiments.

Supplementary Figure 3: Spatial mapping of cell types in the human temporal cortex

Supplementary Figure 4: Spatial mapping of layer-specific cell types.

Supplementary Figure 5: Mapping the cellular compartments using non-negative matrix factorisation (NMF).

Supplementary Figure 6: Mapping the cellular compartments using non-negative matrix factorisation (NMF).

Supplementary Figure 7: UMAP plots of the Visium data

CHAPTER 1: LITERATURE REVIEW

1.1 Introduction

Human neurodevelopment is a complex and tightly regulated process which occurs over an extended period of development (Kang et al., 2011; Wang and Wang, 2019). The human brain is functionally divided into distinct regions composed of a diversity of cell types which have different gene expression profiles (Song et al., 2021; Wang and Wang, 2019). Therefore, it is of utmost importance and advantageous to be able to understand the activity of the brain at a molecular level (Song et al., 2021; Wang and Wang, 2019), especially over the course of brain maturation. This is essential for a comprehensive understanding of how the brain functions during development.

The gene expression levels, in each cell type, in the human brain changes during development and maturation, from embryogenesis until adulthood (Werling et al., 2020). Understanding these changes in gene expression that occur during brain maturation allows scientists to obtain information about the molecular and cellular processes underlying development (Werling et al., 2020). It also allows the quantification of compositional changes in cell types at different developmental stages, and it reveals the dynamics of gene expression within a cell type during development.

However, understanding the cellular and molecular heterogeneity of the human brain continues to be a challenge for several reasons. Firstly, most of the tissue samples that are available are postmortem, thereby limiting the quality of the samples (Darmanis et al., 2015). Additionally, genome wide transcriptome methods commonly used to study cell population such as DNA microarray and bulk RNA seq, only provide an average gene expression across the cells in a tissue (Yoosuf et al., 2020). These methods provide limited information about the heterogeneity within tissues. Single cell RNA sequencing (scRNA-seq) has made it possible to explore cellular compositions and to obtain cell-type specific gene expression information.

A study conducted by Darmanis et al., used scRNA-seq to capture the cell complexity of the adult and fetal brain at a whole transcriptome level (Darmanis et al., 2015). They were able to classify cells into neuronal, glia and vascular cell types in the brain. Using scRNA-seq they observed several changes that occur in neurons between fetal brain and adult brain. For example, their results indicated that adult neurons expressed a high level of *SNAP25*. Another

study conducted by Lake et al., used scRNA-seq on six areas of an adult human brain. Using differential expression analysis, they were able to define 17 neuronal subtypes and found different cellular compositions within distinct regions (Lake et al., 2016). They differentiated 972 inhibitory neurons that generally encompass interneurons from the 2 253 excitatory neurons that generally encompass pyramidal neurons. They found that the inhibitory neuron subtypes distributed spatially from the pial surface to the white matter of the neocortex. This effort to define the cellular complexity of the human brain has been expanded to include several public scRNA-seq databases constructed for the brain, including the Allen Brain Cell Types Database (Song et al., 2021).

While scRNA-seq provides information on gene expression at cell-specific level it does not provide the location of the cell types because it requires the tissue to be dissociated, which disrupts the spatial context of cells in the tissue. Being able to link the molecular cell types to their location is important to fully understand brain function, development, and disease. Spatial information helps in understanding how cell localization in the tissue influences gene expression and how neighboring regions interact at gene expression level during brain maturation. Therefore, a study focusing on spatial information of the different cell types within the different brain regions during brain maturation is in demand. Obtaining information on the location of cells in the human brain is important for understanding the tissue functionality and corresponding pathological changes.

scRNA-seq studies have been able to improve our understanding of how the structure and complexity of the human brain matures. However, these studies mainly focused on postmortem prenatal and adult brain stages, they also lack spatial information. Therefore, very little is known about the spatial information of the human brain tissue during postnatal brain maturation, particularly the juvenile stage. This stage is quite important as the brain goes through many changes due to several influences such as the environment, heredity, and sex hormones. This study aims to contribute to the cell atlas of the developing human brain by using Visium spatial transcriptomics to obtain the spatial information of the ante-mortem brain as it matures post-natally from juvenile to adult state.

1.2 The human brain

The human brain is the most complex organ of the human body. Which consist of complex interconnected net of neurons organized in regions, sub regions, nuclei, and layers. The human brain produces every memory, feeling and experience of the world. It is divided into three basic units namely: hindbrain, forebrain, and midbrain. The hindbrain includes the myelencephalon, which gives rise to the medulla oblongata, and metencephalon, that develops into the pons and the cerebellum (Fig. 1.1). The cerebellum is responsible for coordinating movement. The midbrain controls reflex actions and is involved in the movement of the eye. The forebrain is the largest part of the brain, which is divided into telencephalon, which is also known as the cerebrum (including cerebral cortex, hippocampus, and sub cortical nuclei) (Fig. 1.1), and the diencephalon. This part of the brain is involved in the intellectual activities of the brain.

The cerebral cortex is the coating around the surface of the cerebrum and cerebellum which consist of the outer layer called cortex (gray matter) and the inner layer (white matter) (Fig. 1.1). The gray matter consists of a few myelinated axons and high number of neuronal cell bodies and synapses. The white matter has more lipids and contains less neuronal cell bodies than the gray matter. There are four lobes in the cortex namely, parietal lobe, frontal lobe, occipital lobe, and temporal lobe. The human brain tissue contains different cell types which are divided into neuronal and non-neuronal cells. The non-neuronal cells include glial cells and endothelial cells.

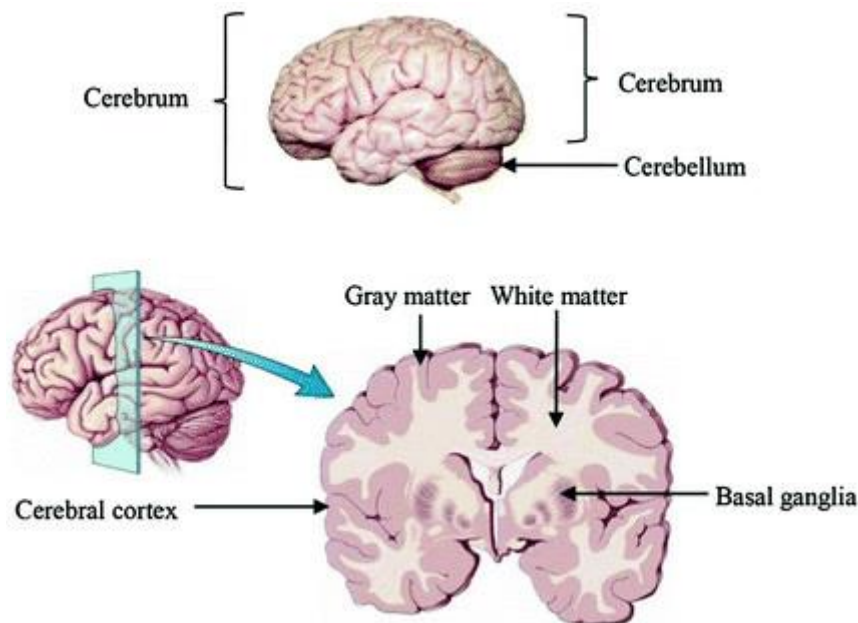


Figure 1. 1 The organization of the brain

(Top) Shows the two major regions of the brain that are visible from the exterior. (Bottom) This shows the cerebral cortex which is the outermost layer of the cerebrum that consist of gray matter, white matter and basal ganglia. Adapted from <http://www.bioon.com/book/biology/whole/image/1/1-6.tif.jpg>, retrieved 20 August 2021)

1.2.1 Neurons

The human central nervous system contains approximately 86.1 billion neurons (Silbereis et al., 2017). Neurons, also called nerve cells, are the main functional units of the nervous system. They are electrically excitable cells that transmit nerve impulses. Neurons consist of three parts namely: cell body, dendrites, and axon (Fig. 1.2). The cell body contains the nucleus, and this is where most of the molecules needed by neurons are synthesized. Dendrites are thin structures that extend from the cell body. Their function is to receive nerve impulses and send them to the cell body. Axon is a long thin extension of the cell body, which functions to transmit nerve impulses from the cell body to other cells. Neuronal cells can be classified according to their morphology, function and their gene expression profiles. Morphological classification includes stellate cells which are small interneurons that have a star shaped dendritic tree. Many neurons are classified based on their functions. For example, the sensory neurons, which are also called afferent neurons, are defined by their role in carrying information from tissues and organs to the central nervous system. On the other hand, motor neurons, which are also called efferent neurons, carry information from the central nervous system to the muscles and glands. Interneurons carry information back and

forth often between sensory and motor neurons within the spinal cord or brain (Neurons and Glial Cells, 2020).

More recently, single-cell transcriptomics has been used to classify neurons according to their gene expression profiles. Hodge and colleagues used single nucleus RNA-sequencing analysis to investigate the different cell types of the middle temporal gyrus of the human cortex (Hodge et al., 2019). In this study they identified 75 transcriptomically distinct cell types: 24 excitatory neuron sub-types that express the vesicular glutamate transporter *SLC17A7*, 6 non-neuronal cell sub-types that express the glutamate transporter *SLC1A3* and 45 inhibitory neuron sub-types that express the GABAergic interneuron marker *GAD1* (Hodge et al., 2019).

Using scRNA-seq Bakken and colleagues investigated the molecular information of the thalamocortical neurons (TC) and how these relate to differences in dorsal lateral geniculate (dLGN) nucleus between human, macaque, and mice (Bakken et al., 2021). To obtain the molecular information they profiled gene expression in single nuclei and cells using RNA-sequencing. In line with previous evidence which states, that there are at least three distinct types of thalamocortical (TC) neurons in the dLGN, Bakken and colleagues have further identified an additional type, revealing a total of four distinct types of TC neurons in the primate dLGN namely: magnocellular (M) neurons, parvocellular (P) neurons, and two types of koniocellular (K) neurons (Bakken et al., 2021).

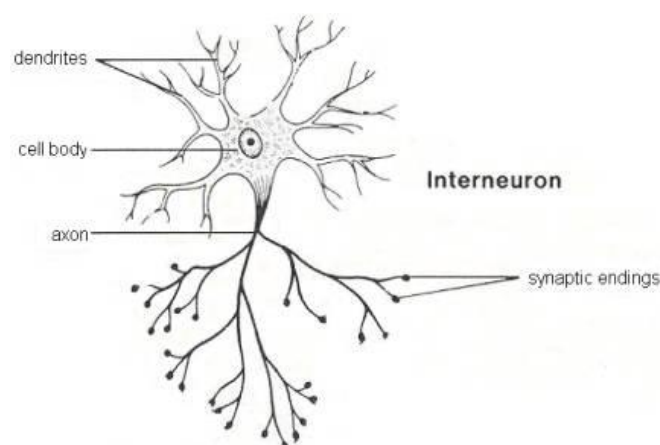


Figure 1. 2: The basic structure of a neuron.

Indicating the three main parts that makes up a neuron namely: dendrites, cell body and axon. Adapted from (<http://www.biologymad.com/nervoussystem/nervoussystemintro.htm>) retrieved on 20 August 2021

1.2.2 Glia cells

Glia cells can be divided into three different classes: oligodendrocytes, microglia, and astrocytes.

1.2.2.1 Oligodendrocytes

Oligodendrocytes are the myelinating cells of the central nervous system (Bradl and Lassmann, 2010). They provide structural support by wrapping neuronal axons in a thick membrane of myelin (Budday et al., 2015; Gonzalez et al., 2017). This enables a rapid flow of electrical signals through neural networks. Oligodendrocytes are generated from oligodendrocyte precursor cells (OPCs) which arise from precursor cells in the ventral part of the ventricular zone and the subventricular zone of the neural tube. After they are generated they migrate to their final destination and differentiate into mature oligodendrocytes (Bradl and Lassmann, 2010; Budday et al., 2015). During embryonic development OPCs are first observed at week 10 and higher numbers start appearing at week 15. At week 20 OPCs differentiate into mature oligodendrocytes which are mostly found in the white matter containing bundles of axons connecting brain regions (Fig. 1.3a).

1.2.2.2 Astrocytes

Astrocytes are the most abundant cell type in the central nervous system (Fig. 1.3c). They have several functions such as the formation of blood brain barrier, response to inflammation and axonal guidance. They play an important role in recycling of glutamate, synaptic transmission and information processing (Gonzalez et al., 2017). Astrocytes can be divided morphologically into protoplasmic and fibrous subtypes (Budday et al., 2015). Protoplasmic astrocytes are in the gray matter, they have irregular branches. Fibrous astrocytes are widely spread in the white matter, and they have regular cylindrical fibers. Astrocytes in humans are larger, more structurally complex and are more diverse than astrocytes in rodents (Budday et al., 2015). Astrocytes develop from radial glial cells after mass production of neurons, the radial glial cells switch to generating astrocytes. This switch indirectly affects the number of astrocytes in the adult brain (Budday et al., 2015). (Fig. 1.3c)

1.2.2.3 Microglia

Microglia are the only cells that are not differentiated from neural stem cells, instead they derive from myeloid progenitor cells migrating through blood vessels. They migrate into the brain during embryonic development. They are found in both the gray and white matter.

Microglia in the white matter have an elongated cell nucleus and align their fibers with neighbouring axons whilst the ones in the gray matter have a small, dense nucleus with fine branches extending in multiple directions (Fig. 1.3b). They play an important role in performing vital immune functions. Microglia are associated with various neurological disorders (Gonzalez et al., 2017).

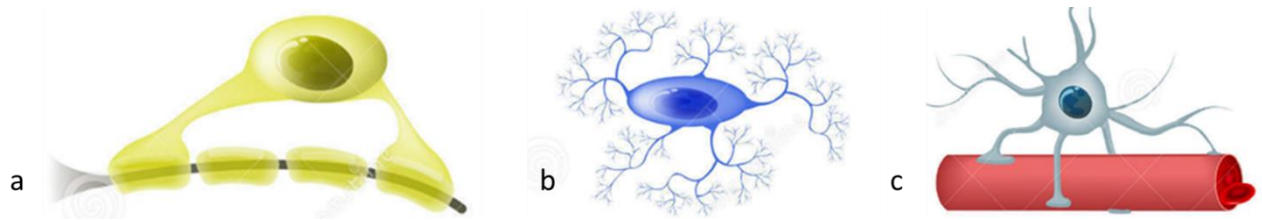


Figure 1. 3: Three different classes of glial cells.

(a) Oligodendrocytes, (b) microglia and (c) astrocytes. (Adapted from <https://www.dreamstime.com/stock-photo-supporting-cells-neuroglia-glia-cells-types-classification-microglia-astrocytes-oligodendrocytes-schwann-ependymal-image47597814>)

1.3 Different stages of brain development.

Human brain development is a protracted process that begins in the third gestational week and in some regions of the brain the development continues into the mid-20s (Fig. 1.4). During prenatal months, brain development is mostly under genetic control, however the environment also plays a role. For example, a lack of nutrition can negatively influence the development of the brain. The first brain structure is called the neural tube (Fig. 1.4). This structure forms during the third week of gestation. The process of the formation of the neural tube is called neurulation.

Brain development in the fetal period starts from the ninth week until the end of gestation (Stiles and Jernigan, 2010). Several anatomical brain changes occur during this period. These anatomical changes reflect the changes occurring at the cellular level. Fetal brain development centres on the processes of neuron production, migration, and differentiation. The production of neurons begins in the embryonic period all the way through mid-gestation. Once the neurons are produced, they migrate away from the proliferative regions of the ventricular zone (Fig. 1.5). Neurons will form the neocortex with cells migrating in a coordinated manner forming the six-layered neocortical mantle (Fig. 1.5). These neurons differentiate into mature neurons (Fig. 1.2) with axons and dendrites producing

neurotransmitters and neurotrophic factors. Axons and dendrites form the fibre pathways of the brain neural networks.

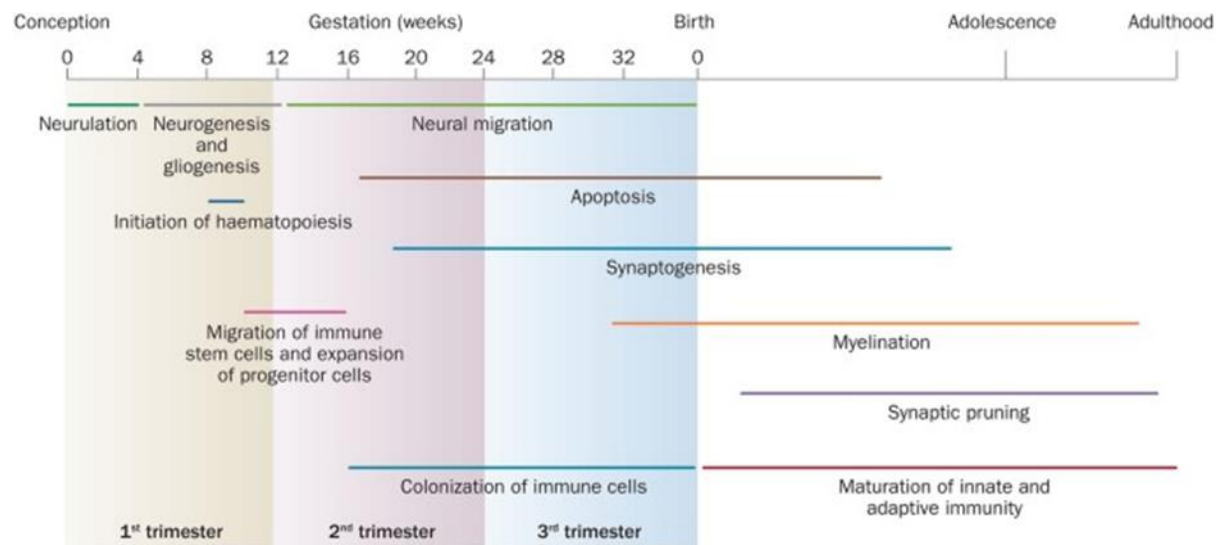


Figure 1. 4: Different stages of brain maturation.

Neurulation is the first stage of brain development, followed by neurogenesis and neural migration. Myelination occurs before birth all the way through adolescence whilst synaptic pruning occurs from birth through to adulthood. Adapted from (Knuesel et al., 2014)

Most key aspects of neural development in humans occur prenatally however the brain continues to develop postnatally (Knuesel et al., 2014). Events such as the production and migration of neurons occurs prenatally whilst proliferation and migration of glial progenitors continues for an extended period after birth. Several other aspects of brain development continue postnatally such as sculpting of synaptic connections, neurotransmitter function, synaptic pruning, and myelination (Fig. 1.4).

The formation of synapse, known as synaptogenesis, begins around 23 weeks of gestation but majority of this process occurs postnatally (Molliver et al., 1973). The time course of synaptogenesis varies in different regions of the brain. In the primary visual cortex, synapse formation peaks between 4 and 12 months, while in the prefrontal cortex, it reaches its maximum level between 8 months and the second year of life, following a similar timeline but with a delay compared to the visual cortex (Tau and Peterson, 2010). Synapses are important components of neurons, that allow information to flow across the nervous system. This helps the body to adjust in response to environmental stimuli and controls the regulation of body functions, memories, and emotions. Therefore, optimal synaptic communication is crucial for ensuring normal brain function, as even minor disruptions in synapse performance can

contribute to the development of brain disorders (Lepeta et al., 2016). Furthermore, numerous genes related to synaptogenesis are often linked to neurological or psychiatric disorders, indicating a significant connection between synaptogenesis, brain functioning, and the development of pathological conditions (Qi et al., 2022).

Synaptic pruning is a natural process in brain development that helps create and strengthen important neural connections by getting rid of unused synapses while keeping the active ones (Germann et al., 2021). Although pruning happens throughout life, there are critical developmental periods, like adolescence, when it reaches its peak. During this time, brain regions responsible for advanced thinking and cognitive abilities undergo significant pruning. Disturbances in synaptic pruning can result in severe alterations in neural circuitry, which may underlie the neuropathological changes observed in various neurological and neuropsychiatric disorders (Cardozo et al., 2019; Sekar et al., 2016; Yilmaz et al., 2021).

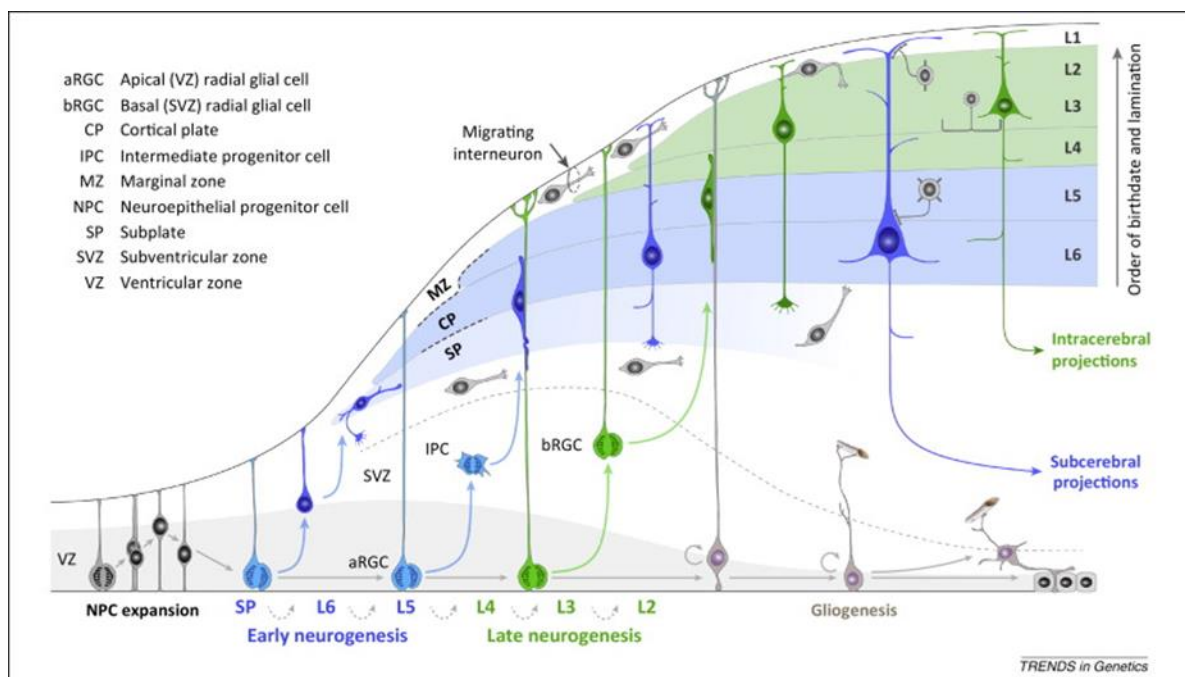


Figure 1. 5: Schematic showing neurogenesis in the mammalian cortex.

The diagram indicates the neocortical neurons maturation into cortical projection neurons (CPNs), which show layer- and subtype-specific morphology and axonal projection patterns. The newly born neuronal cells migrate vertically from the Ventricular zone (VZ) to the cortical plate (CP). The youngest neurons migrate to form the preplate. Later migrating neurons splits the preplate forming the marginal zone and subplate. The early born neurons settle in L5-6 (early neurogenesis) whilst the later-born settles in layers 2 through 4 (late neurogenesis). Adapted from (Shibata et al., 2015)

1.4 Brain development is under genetic control.

Brain function and development is strictly controlled by the regulation of genes across its anatomically distinct structures. Each cell in the brain shows a different pattern of gene expression. The transcriptome is the total set of RNA transcripts present in the cell. Investigating cellular transcriptomes and how they change both temporally and spatially in different cell types will assist in understanding the development of each cell, their molecular mechanisms and how the changes in the transcriptional activity may contribute to disease. Several technologies are available to study an organism's transcriptome.

1.5 Previous technologies used to obtain gene expression data.

The first transcriptomic method developed in the quest to analyse the cellular transcriptome was Serial Analysis of Gene Expression (SAGE) (Velculescu et al., 1995; Lowe et al., 2017). In SAGE, random transcript fragments are sequenced using Sanger sequencing and are quantified by matching them to known genes. In 2001, SAGE together with RT-PCR was used to identify RNA transcripts which were differentially expressed in the frontal cortex of individuals with bipolar disorder compared with other psychiatric and control conditions (Sun et al., 2001). In this study they found that the levels of RNA transcripts encoding the serotonin transporter protein and components of the NF-kappa B transcription factor complex were higher in individuals with bipolar disorder compared with unaffected controls (Sun et al., 2001). However, high throughput methods that sequenced the entire transcriptome, providing additional information about the structure of the transcriptome quickly took over from the SAGE approach (Lowe et al., 2017).

In the mid-1990s and 2000s microarrays and bulk RNA-seq were developed. The microarray technique uses the ability of a fluorescently labelled cDNA to bind to a specific DNA template (Fodor et al., 1991; Schena et al., 1995). These DNA arrays are microscope slides that contain thousands of DNA probes representing genes in defined positions. Often these slides are referred to as gene chips or DNA chips. The amount of bound cDNA is calculated with microarray software. This calculation shows the abundance of the genes being expressed. Evans and colleagues used DNA Microarrays to investigate the transcriptional profiles of two divergent brain regions, namely cerebellar cortex and cerebral cortex, as well as two closely related structures namely the anterior cingulate cortex and the dorsolateral prefrontal cortex

(Evans et al., 2003). This study was replicated across three independent laboratories to address false-positive and false-negative results using microarray technology. They discovered that more than three thousand transcripts were differentially expressed between the cerebellum and cerebral cortex. Approximately 1600 transcripts were reproducible across all three laboratories. One laboratory found a few (559) transcripts that were differentially expressed between the two neocortical regions (anterior cingulate cortex and the dorsolateral prefrontal cortex) and of those only four were reproducible across all laboratories. The few transcripts that were found to be differentially expressed between the two neocortical regions supports the hypothesis that these two neocortical regions are closely related. However, there might be more differences between these neocortical regions that were missed because of the inefficiencies of Gene Chips in reliably detecting relatively low-magnitude changes (Evans et al., 2003). Another shortfall of gene chips used in microarray is the inconsistency in detecting low abundance transcripts (Evans et al., 2002). This is indicated in this study by the differences in the number of transcripts found across the independent laboratories.

Bulk RNA-seq uses the principle of sequencing all the transcript cDNAs in abundance using high-throughput sequencing technologies (Fig. 1.6). It differs from the other high throughput sequencing methods because it uses transcript enrichment, fragmentation, amplification and single- or paired-end Next Generation Sequencing. In bulk RNA-seq analysis, the expression value of one gene from a sample represents the mean of all expression values of the bulk population of the cells. Therefore it is not sufficient to use bulk RNA-seq for biological research which seeks to detect differential gene expression between the different cell types that make up a tissue sample (Wang et al., 2019).

This is indicated by a study conducted by Werling et al., which used bulk RNA-seq data to detect changes in gene expression in the dorsolateral prefrontal cortex (DLPFC) as this brain region develops from embryonic stages to young adulthood (Werling et al., 2020). They used their bulk RNA-seq data to look at the relationship between the change in proportion of different cell types and gene expression profiles by assessing the trajectories of specific genes that are known markers for ten distinct prenatal and adult cortical cell types. Previous studies have shown that developmental age is the greatest source of between sample variance in gene expression especially during the “late-fetal transition”, which is a period between 22

post-conception weeks and 6 postnatal months. They found that the estimated profiles of all ten cell types vary dramatically during the transition from the fetal period to infancy. They showed that the proportions of radial glia and fetal neurons decrease as mature neurons and glial cells increase. The authors suggest that their results support the hypothesis that varying cell type proportions are major contributors to the late-fetal transition in the DLPFC. However, they also mentioned that distinguishing cellular composition effects from differential expression in a cell type requires further analysis at single cell level (Werling et al., 2020).

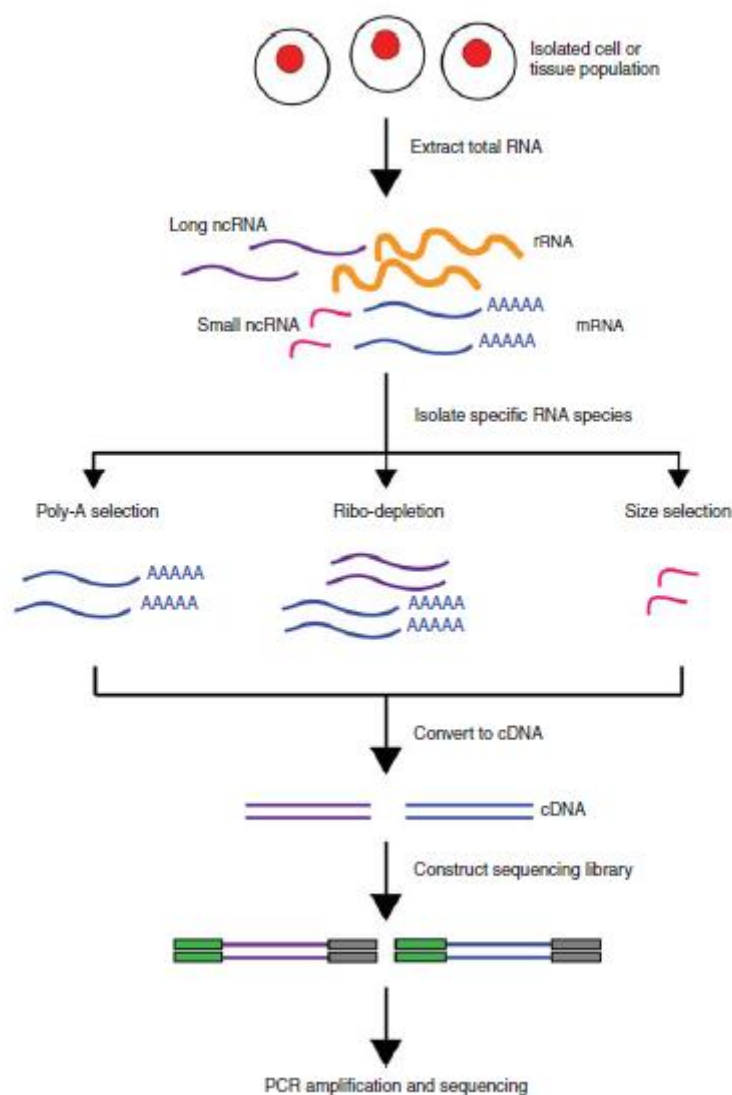


Figure 1. 6: The process of bulk RNA sequencing.

The process involves the isolation of RNA from the tissue of interest, the conversion of RNA to cDNA followed by sequencing and finally analyzing the data. Adapted from (Kukurba and Montgomery, 2015)

1.6 Using single cell and single nucleus RNA sequencing to obtain cell-type specific gene expression data.

The recent advances in high-throughput DNA technologies has made it possible to obtain gene expression information at single-cell level (Ecker et al., 2017a; Hwang et al., 2018; Trapnell and Liu, 2016; J. Wang et al., 2020). The first report on studying single cells' gene activity using next generation sequencing was published in 2009 (Tang et al., 2009). This study focused on understanding how cells change and develop during the early stages of growth in the mouse embryo (Tang et al., 2009). Over the past few years, a high-throughput technology referred to as single cell or single nucleus RNA sequencing (sc/snRNA-seq) has provided new information on the complexity of brain tissues based on genes expressed by individual cells (Fig. 1.7) (Cao et al., 2019; Ecker et al., 2017b; J. Wang et al., 2020). In previous studies, before scRNA-seq was available, neuronal cell types were characterized and classified using parameters such as anatomical location, expression of certain marker genes, cell morphology and patterns of connectivity (Ecker et al., 2017a). However, they were unable to provide unbiased information on the cell type-specific gene expression. scRNA-seq is efficient for cataloging cell types, identifying cell type-specific marker genes and discovering novel subtypes (Ecker et al., 2017a).

There are several strategies for cell sorting and capture in snRNA-seq such as microfluidics devices, which include Fluidigm's C1 and "ultra-high throughput" droplet-based technologies. The droplet-based approaches, including the commercial 10X Genomics platform and DropSeq, use microfluidic chips and single beads in oil encapsulated droplets to isolate single cells or nuclei (AlJanahi et al., 2018). The single beads are coated in oligonucleotides that are made up of a poly (T) tail at the 3' end, a cell/nucleus barcode at the 5' end and a library of unique molecular identifier barcodes. The unique molecular identifier is used to directly quantify the expression level of each gene, the poly(T) tail at the 3' end is used to capture the cellular mRNAs and the cell barcode is used to record the transcript's cell/nucleus of origin.

Droplet-based techniques have gained popularity because they require less starting material. The main drawback of droplet-based methods that use short-read sequencing approaches is the molecular barcode, which limits the sequence analysis to short stretches at the mRNA 3'

end therefore cannot give information on the isoforms of mRNA. (Ecker et al., 2017b). Chromium 10X Genomics has gained attraction because it allows genome-wide expression profiling for thousands of cells at once and it is a high throughput processing method which allows for analysis of rare cell types in a sufficiently heterogeneous biological space (Wang et al., 2019). However, clinical samples must be handled with caution to prevent disturbance of the existing cellular characteristics (Hwang et al., 2018).

The process of scRNA-seq typically involves preparation of the single cell or single nucleus library using the droplet- based techniques, followed by sequencing of the library using compatible sequencing platforms such as the illumina and MGI sequencer. Using a set of analysis pipelines the sequencing data is changed into a FASTQ files followed by read alignment to a genome. This aligned data is further analyzed using a range of bioinformatics tools to sort nuclei into clusters based on similar gene expression profiles.

The major advancement achieved by scRNA-seq thus far has allowed research to go beyond descriptive studies of cell states, such as obtaining information on tumor heterogeneity. Such findings have made this technology exciting for medical applications. For example, it can be used for illuminating tumor features which cannot be determined from conventional bulk transcriptomic studies (Hwang et al., 2018). Another potential application of single-cell technologies is identifying genes that are involved in stem cell regulatory networks (Hwang et al., 2018). In addition, several projects are available to help us understand the brain.

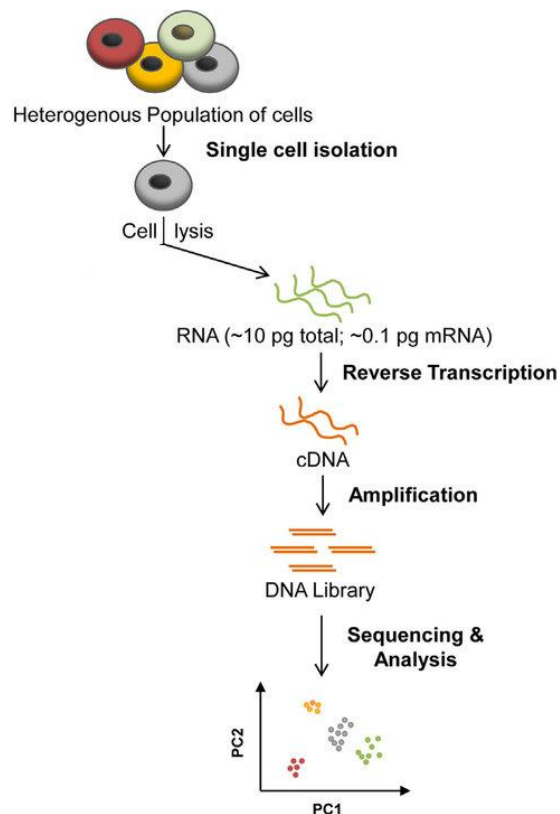


Figure 1. 7: The process of single-cell RNA sequencing.

The process involves the isolation of single-cell-barcoded cDNA from the tissue followed by sequencing and finally analyzing the data.

(Adapted from https://www.researchgate.net/publication/309660136_Exploring_Viral_Infection_using_Single-Cell_Sequencing/figures?lo=1)

1.6.1 Recent initiatives to understand brain development.

A number of new initiatives using new technologies, such as scRNA-seq, have been developed in order to help us understand the genetic mechanisms of brain development and functioning (L. Liu et al., 2020). The Allen Institute for Brain Science (AIBS) started in 2003. This initiative aims to expand our understanding of the biological and biophysical properties of the cerebral cortex. Importantly, this initiative aims to create a “Periodic Table of Cell Types”. This involves the use of a combination of single-cell transcriptional, morphological, electrophysiological, connectional, and functional investigations to derive an ontology of neuronal cell types in mouse and human cortex. The Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) initiative launched in 2013. The ultimate goal of BRAIN is to understand how the brain and body create our thoughts, motivations and how to help those whose brains are failing (Grillner et al., 2017). In the same year, The European Union’s Human

Brain Project (HBP) was started. The HBP aims to understand the simulations and modelling of the brains of mice and humans, based on a detailed neurobiological knowledge of different parts of the brains. This was shortly followed by Japan Brain/MINDS in 2014. This initiative aims to map the brain of a small New World monkey, the common marmoset, to gain a better understanding of the primate brain and develop knowledge-based strategies for the diagnosis and treatment of human psychiatric and neurological disorders (Grillner et al., 2017). In 2016, Human Cell Atlas (HCA) was started. This global initiative aims to map every cell type in the healthy body from development to adulthood by state-of-the-art single cell and spatial genomics, along with advanced computational techniques (Regev et al., 2017).

Included in this effort are a number of scRNA-seq studies that have been focusing on deepening our understanding of cells in the brain and the relationships among brain cells during development (Carter et al., 2018; Darmanis et al., 2015; Fan et al., 2018; Hwang et al., 2018; Kang et al., 2011).

1.6.2 Previous studies that used scRNA-seq to investigate gene expression in the brain.

Fan et al. used scRNA-seq on different brain regions (20 cortical regions, as well as the medulla and pons) from human mid-gestation embryos. They identified multiple subpopulations of cells (microglial, neurons & astrocytes) in each region and analyzed changes in gene expression and neuronal maturation in these cell sub-types (Fan et al., 2018; Tang et al., 2019). By comparing their neuronal embryonic data to published scRNA-seq adult neuronal data (Lake et al., 2016) they observed drastic differences between embryonic neurons and adult neurons. They found that the maturation degree of embryonic neurons is different from the adult neurons. For example, the expression level of upper layer 2/3/4 marker *CUX2* was higher for mature neurons (Fan et al., 2018). Another study, conducted by Darmanis et al., used scRNA-seq to capture the cell complexity of the adult and fetal brain at a whole transcriptome level (Darmanis et al., 2015). They were able to classify cells into neuronal, glia and vascular cell types in the brain. Using scRNA-seq they observed several changes that occur in neurons between fetal brain and adult brain as well as observed genes that display an expression gradient that reflects the transition between replicating and quiescent fetal neuronal populations. This analysis identified three groups of cells, including adult neurons that express high levels of *SNAP25* and *GAD1*, actively dividing fetal neuronal

progenitors (MKI67+ and PAX6+), and quiescent subplate fetal neurons (MKI67–, DCX+, and TUBB3+). (Darmanis et al., 2015).

Eze et al., used scRNA-seq to characterize the earliest stages of human brain development in different regions of the developing human brain (telencephalon, diencephalon, midbrain, hindbrain and cerebellum) (Eze et al., 2021). They sequenced cells from 10 individuals during the first trimester of human development (Carnegie stages (CS) 12 to 22). At CS13 they observed that characteristic transcription factors such as HOX genes, *PAX7*, *GBX2*, *NKX2-1*, and *FOXG1* respectively segregated these regions: hindbrain, midbrain, thalamus, medial ganglionic eminence and cerebral cortex from one another (Eze et al., 2021).

To better understand the early stages of cortical development, they performed scRNA-seq analysis on the telencephalon. They identified 63 clusters which they annotated as neuroepithelial cells, radial glia, intermediate progenitor cells (IPCs), neurons or mesenchymal-like cells. Using the gene *SOX-2* they were able to differentiate radial glia from neuroepithelial cells. *SOX-2*-positive progenitor clusters that contained evidence of neurogenic genes were labeled as radial glia, whilst *SOX-2*-positive progenitors that did not contain neurogenic genes were labeled neuroepithelial cells. They further investigated the heterogeneity of the neuroepithelial and radial glia progenitor populations by sub-clustering the progenitors. The resulting analysis yielded nine subclusters of progenitor populations which were all marked by a strong *VIM* and *SOX2* expression (Eze et al., 2021). These results importantly indicated more heterogeneity in the progenitor population than previously shown.

The information from these studies contributes to the cell atlas of a human brain, and can be used to identify novel pathways which are involved in neuro-related diseases (Tang et al., 2019) as well as capture cell type-specific molecular signatures related to diseases (Ecker et al., 2017a). To understand how the gene expression of the human brain responds to a disease, it is important to understand the molecular profile of the antemortem human brain. It is also important to have this information from a range of ages to determine how the cell-specific gene expression of the brain changes as it matures.

1.6.4 Limitations of scRNA-seq

Just like any other technology, scRNA-seq has disadvantages. For example, it results in a high number of dropouts. Dropouts occur when some transcripts appear to be highly expressed in one cell but are missed in another cell (Chen et al., 2019; Wang and Wang, 2019). This is caused by low starting material in scRNA-seq and could also be caused by failed amplification of the original cDNA. In comparison to bulk RNA-seq, scRNA-seq produces a higher technical noise and more biological variation which increases the number of challenges for computational analysis of scRNA-seq data (Chen et al., 2019).

A major challenge for scRNA-seq is the loss of spatial information of individual cells due to enzymatic or mechanical tissue manipulation during the process of tissue dissociation (Ren et al., 2020; Roth et al., 2020; Teves and Won, 2020). Tissue dissociation disrupts the laminar organization of the cells within the cerebral cortex. Getting information on the laminar position of a cell) is important because the layered cortical network provides the basis to identify the underlying connectivity rules and organizing principles by which the brain integrates cortical activity (Fig. 1.8. To overcome this limitation spatial information can be obtained by integrating imaging and spatial barcoding using spatial transcriptomics.

1.7 Spatial transcriptomics

1.7.1 What is spatial transcriptomics?

Spatial transcriptomics approaches are groundbreaking molecular technologies that allow scientists to analyse the gene activity in a tissue sample by providing an unbiased map of RNA targets throughout the tissue sample (Pierre, 2021; 10X Genomics). Spatial transcriptomics has been used in understanding the human and mouse brain spatial atlases, discovering spatial heterogeneity in disease and delineating their spatial blueprints. It can be used to study cell-to-cell communication because cells in a close place such as the tissue are more likely to interact with each other.

1.7.2 Why is spatial transcriptomics important?

Spatial transcriptomics allows researchers to study the relationship between cells and their organization within a tissue, which is critical for understanding both normal development and disease pathology. This is particularly important in the context of the brain. The functioning of the brain is related to how the cells are organized and where they are located within the

brain (Maynard et al., 2021). Communication between cells and their surroundings allows for functional organization of the cells in the brain. This relationship can be seen in how the cells residing in the human cerebral cortex are arranged into six cortical layers. The cells in each layer show distinct gene expression patterns and exhibit differing patterns of morphology, physiology, and connectivity (Fig. 1.8) (Maynard et al., 2021). Obtaining the spatial organization of gene expression in the brain will help us to understand how the cell localization in the tissue influences its gene expression profile and how neighboring regions interact at gene expression level during brain maturation.

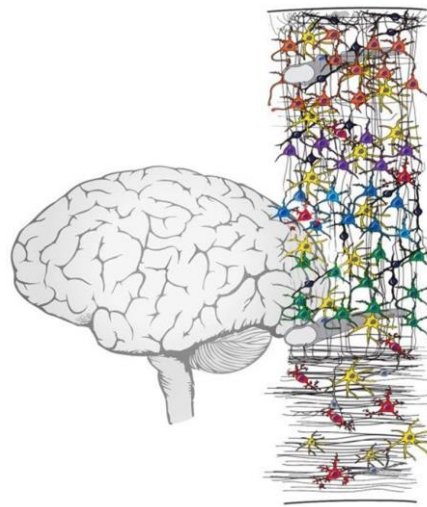


Figure 1. 8: Laminar organization of the cortex.

The diagram shows the different layers of the cortex made up of different types of cells. The cells are in different colours. Adapted from Piero Carninci, 2018

1.7.3 Technologies in spatial transcriptomics

Spatial transcriptomics is used for spatial molecular profiling to visualize thousands or millions of cells and transcripts within the context of intact tissue. In the past positional information for gene expression was obtained using a microdissection approach, which involved physically isolating regions of interest from within a tissue sample and analyzing their gene expression profiles separately (Asp et al., 2020). However, it is difficult to cover a large area with microdissection-based approaches. With microdissection approaches the regions of interest need to be known beforehand since the analyses can only be carried out on the collected samples. Therefore, information may be missed within the samples. Spatial transcriptomics, on the other hand, allows for the measurement of all the gene activity in intact tissue. The

realization of the importance of spatial organization and the need to preserve the exact position of molecular features has driven technological advancements in the spatial transcriptomics field. Currently spatial transcriptomics methods can be classified into two categories: 1) single-molecule fluorescence in situ hybridization (smFISH) based techniques and 2) barcode-based approaches (Hu et al., 2021).

1.7.3.1 Single-molecule fluorescence in situ hybridization (smFISH)

smFISH was first developed in 1982 (Fig. 1.9) and has further been developed over the years. These techniques measure the expression level of hundreds of genes by using fluorescent labelled nucleotide probes to detect specific mRNAs at different spatial coordinates in a cell or tissue (Hu et al., 2020; Lein et al., 2017; Teves and Won, 2020) (Fig. 1.10). smFISH uses short oligonucleotides probes to target mRNAs and it provides the cell-to-cell variation of many transcripts (Femino et al., 1998; Raj et al., 2008). It provides precise microscopic localization of genes within their cell or tissue environment. It has high sensitivity and spatial resolution however it has a low signal intensity thus requiring high powered microscopy for detection and can also suffer from high non-specific probe binding caused by stray probes. Another drawback of this method is that the genes of interest must be known beforehand, therefore limiting the possibilities of new discoveries.

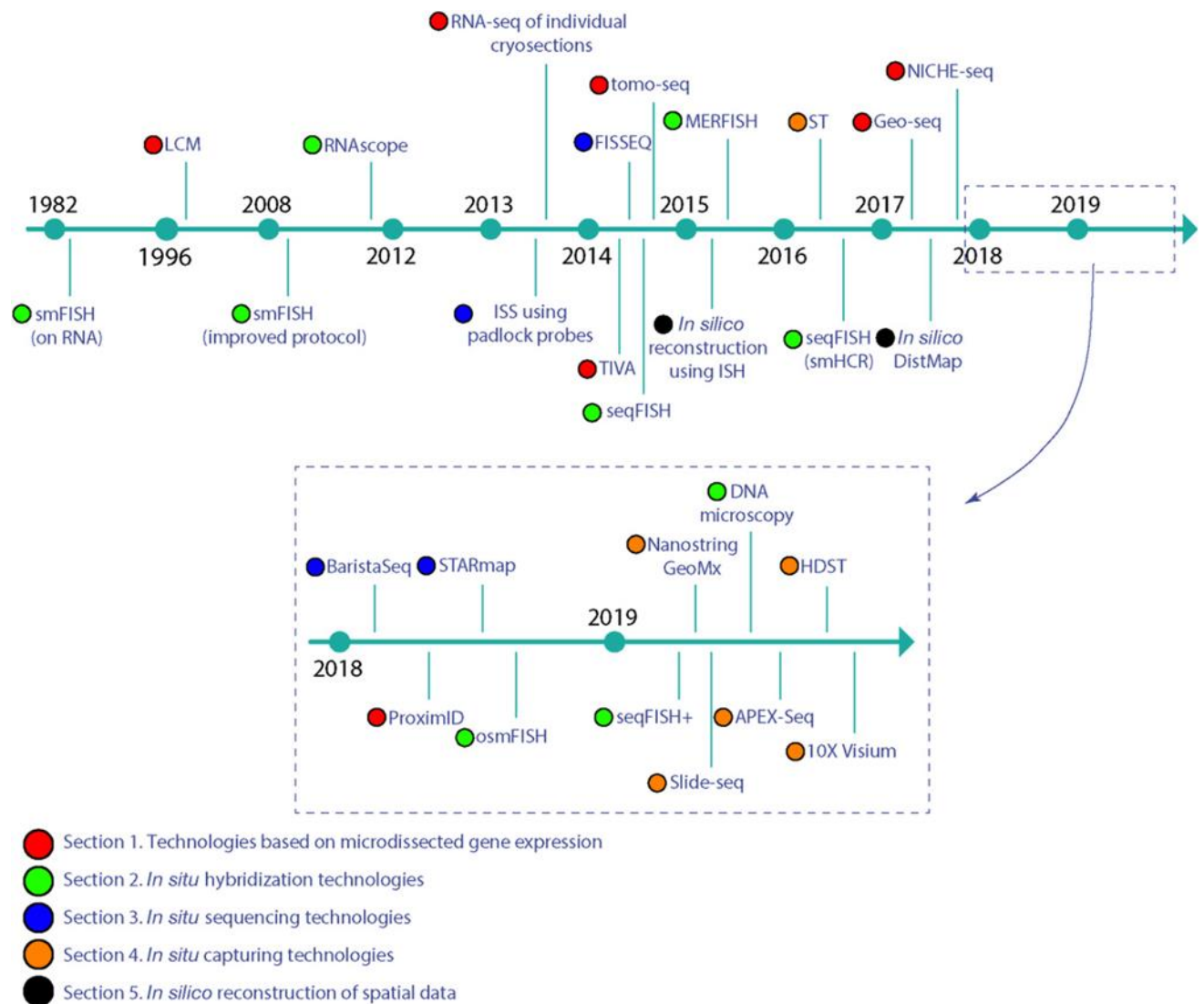


Figure 1. 9: Spatially resolved transcriptomic technologies.

Timeline showing all the spatial transcriptomics technologies used over the years. Adapted from (Asp et al., 2020)

Jolly and colleagues used smFISH on frozen human hippocampal samples (Jolly et al., 2019). In this study they analysed the expression of three cell-type specific genes (*MAP2* for neurons, *SLC1A2* for astrocytes and *P2RY12* for microglia). Their results showed that smFISH is suitable to classify RNA expression in a cell-type specific manner. In addition, they used smFISH to detect three cell-type specific genes known to be up regulated and downregulated in Alzheimer's disease (AD). They used *TREM2* which is mainly expressed in microglia. The polymorphism of this gene is associated with AD. *SNAP25* is expressed in neurons and it is decreased in AD. *DKK1* is expressed in neurons and astrocytes, it is upregulated in AD. In

healthy individuals *SNAP25* is highly expressed and *DKK1* is expressed at low levels. With smFISH they were able to validate that *TREM2* expression is restricted to microglia, *SNAP25* to neurons and *DKK1* to neurons and astrocytes (Jolly et al., 2019). They also observed that *SNAP25* is abundantly expressed whereas *DKK1* is sparsely expressed. This study points out the limitation of smFISH, which is that the genes of interest must be known beforehand therefore smFISH limits the discovery of new possibilities.

Additional technologies were developed to overcome the inefficiency of smFISH, which include a cyclic approach to single-molecule fluorescence *in situ* hybridization method called ouroboros smFISH (osmFISH) (Codeluppi et al., 2018) (Fig. 1.9). osmFISH involves multiple smFISH rounds which are repeated to increase the number of RNA species detected (Fig. 1.10). In each round of osmFISH the detected probes are removed by formamide melting to prepare for the next round. This method has a short hybridization time and a lower background due to tissue clearing. This leads to a high signal-to-noise ratio which improves the sensitivity of RNA detection. It can cover a large tissue area but has a lower multiplex capacity.

Sequential FISH (seqFISH) allows detection of mRNA through multiple rounds of hybridization which allows many molecules to be multiplexed (Eng et al., 2019; Teves and Won, 2020) (Fig. 1.10). seqFISH uses DNase 1 to strip of the hybridized probes in preparation for the next round (Fig. 1.10). For each round of mRNA hybridization, the same probe sequence is used but labeled with a different fluorophore. The disadvantage of seqFISH is the optical density of each targeted transcript which limits the amount of mRNA detected (Teves and Won, 2020). Shah et al., have used seqFISH to detect different transcriptional states in mouse hippocampus. Contrary to previous studies they were able to identify distinct layers in the dentate gyrus corresponding to the granule cell layer and subgranular zone (Shah et al., 2016b). SeqFISH+ is an improved version which has made it possible to detect, with high accuracy, 10 000 genes in a single cell (Eng et al., 2019). SeqFISH+ solved the optical density problem by using pseudo-colours to dilute the cellular RNAs and enabling spatial reconstructions at high resolution (Fig. 1.10).

Multiplexed error-robust FISH (MERFISH) (Chen et al., 2015) is a single-molecule imaging method that allows spatial localizations of thousands of RNA species to be determined in single cells. It does this by using combinatorial labeling, instead of pseudo-colours, and

sequential imaging with error-robust schemes to increase the number of RNA molecules that can be simultaneously imaged (Fig. 1.10) (Teves and Won, 2020; Wang et al., 2018). Wang et al., used MERFISH to conduct quantitative, high-resolution spatial profiling of ~4,200 RNA species within mouse hippocampal neurons to form a cell atlas of the distinct transcriptomic regions inside individual neurons (Wang et al., 2020). In this study they were able to detect approximately 11,000 RNA molecules per cell and found that cell bodies, axons, dendrites, and even smaller dendritic regions, had their own distinct transcriptional profiles. Another study by Fang et al., used MERFISH on 4000 genes in the human brain tissue (middle and superior temporal gyrus). They identified more than 100 neuronal and non-neuronal cell populations and spatially mapped them in the human MTG and STG (Fang et al., 2022). However, seqFISH, seqFISH+ and MERFISH are still limited in their throughput and the ability to accurately measure the level of gene expression. Due to the inherent limitation of spectral overlaps in standard microscopy smFISH approaches can only target a limited number of genes at a time (Asp et al., 2020).

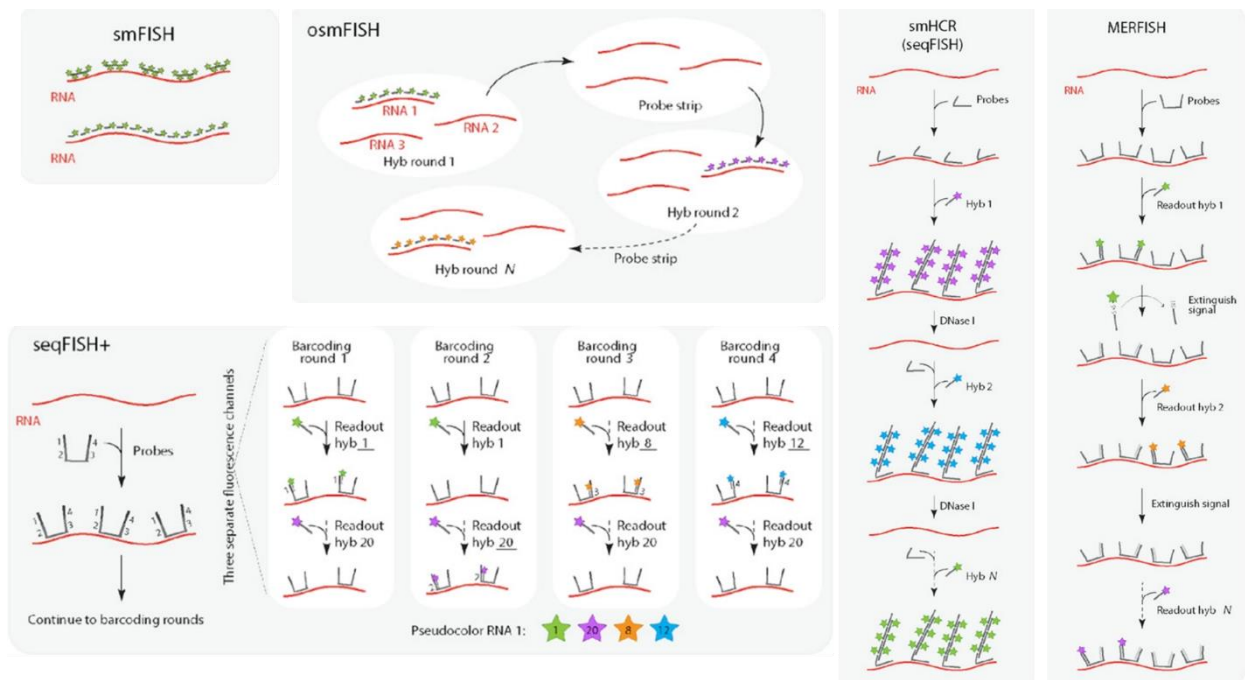


Figure 1. 10: An overview of single-molecule fluorescence in situ hybridization (smFISH) based technologies.

smFISH: In the original smFISH approach, a set of five probes labeled with five fluorophores each are hybridized to individual targets. seqFISH: smFISH probes are hybridized consecutively to the target and stripped off by DNase I in between each round. The same probe sequences are used in different rounds of hybridization, but probes are connected to different fluorophores. smHCR: seqFISH in combination with single-molecule hybridization chain reaction (smHCR) to amplify the signal. MERFISH: A probe set with unique flanking read out regions is used for individual targets. osmFISH: smFISH probes are hybridized in several rounds, along probe removal in between rounds. A few targets are detected per round. seqFISH+: A set of 24 primary probes are hybridized per target, each probe containing four unique flanking read out regions corresponding to four barcoding rounds. Adapted from (Asp et al., 2020)

1.7.3.2 Barcode-based approaches

Barcode-based approaches for spatial transcriptomics involves capturing the mRNA transcripts *in situ*, tagging the mRNA with a spatial barcode and sequencing of the captured transcripts *ex situ*. The tissue of interest is collected on a patterned microarray carrying spatially barcoded oligo dT primers that capture the polyadenylated mRNA molecules released from the tissue section during tissue lysis. This allows scientists to map the mRNA molecules back to their location in the tissue. These approaches provide a high spatial resolution and avoid the typical limitations of direct visualization of mRNA transcripts using microscopy.

In 2016, Spatial Transcriptomics (ST) technology was developed which involves the combination of histology and next generation sequencing to detect and visualize the RNA

molecules in a tissue section with a resolution 100 μm (Ståhl et al., 2016) (Fig. 1.9). This was achieved by placing thin sections of the tissue onto a glass slide which contains about 200 million barcoded RT primers. The primers capture the entire polyadenylated transcriptome in the tissue in barcoded capture regions that have a diameter of 100 μm and a center-to-center distance of 200 μm , over an area of 6.2 mm by 6.6mm (Fig. 1.11). The tissue is first fixed, stained using Hematoxylin & Eosin stain (H&E) to assess the phenotype of the tissue, imaged and permeabilized to release the mRNA molecules to be captured by the barcoded primers. cDNA synthesis is then performed. After cDNA synthesis the tissue is removed and mRNA-cDNA hybrids are released from the array and prepared for sequencing (Asp et al., 2020; Berglund et al., 2020). The sequenced transcript can be mapped to the original location of the extracted tissue using the barcodes.

Using ST, Chen et al. investigated the transcriptional changes occurring in amyloid plaques AD in a mouse model (Chen et al., 2020). ST showed that the multicellular gene co-expression network of plaque-induced genes (PIGs) is prominent in the later phase of AD. The PIGs refer to the communication between astrocytes and microglia that involves other cells, resulting in changes in both the classical complement system and endosomal/lysosomal pathways. These findings link together several pathways that have previously been suggested to play a role in Alzheimer's disease. The researchers also observed a dynamic response of oligodendrocyte genes (OLIG), mainly related to myelin-related genes, and influenced by gradual amyloid buildup. However, unlike the consistent PIGs response, the OLIG response varies in different regions of the brain (Chen et al., 2020).

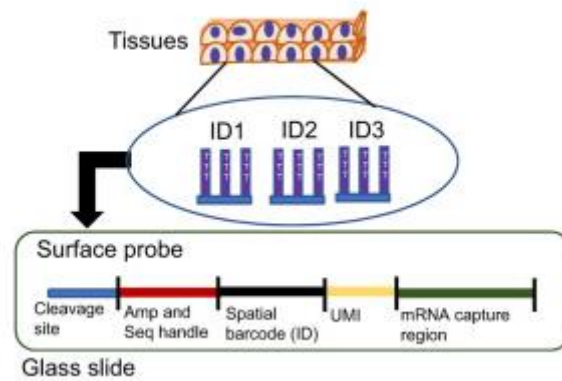


Figure 1. 11: An overview of the glass slide used in ST technology where the tissue of interest binds.

The surface probes, arrayed across the slide, contain a cleavage site, a T7 amplification and sequencing handle, a spatial barcode, a unique molecular identifier (UMI) and a mRNA capture region. Adapted from (Li and Wang, 2021)

Slide-seq technology was developed in 2019. This approach utilizes a glass slide covered in DNA-barcoded beads with known positions to capture mRNAs from the tissue placed on top (Rodriques et al., 2019) (Fig. 1.9). The glass slides contain 10 μm -wide barcoded beads; therefore, it has a spatial resolution of 10 μm (Rodriques et al., 2019). To achieve this, they used uniquely DNA-barcoded 10 μm beads like those used in the Drop-seq approach for scRNA-seq, which they packed onto a rubber-coated glass coverslip to form a monolayer called a “puck” (Fig. 1.12). Slide-seq allows spatial gene expression analysis of frozen tissue, has a high spatial resolution due to the near single-cell resolution and can be easily integrated with large scale single cell sequencing data sets. However, it requires manual sequential hybridization or sequencing by oligonucleotide ligation and detection chemistry (SOLiD) to sequence the DNA barcode bead array, which requires long and repeated imaging processes. It is also difficult to visualize the spatial expression of individual genes, even when collective gene sets can locate the main types of cells. Since the capture spot is 10 μm -wide, very few genes can be captured per spot (Y. Zhou et al., 2020).

Slide-seq was used to localize cell types that were identified by scRNA-seq datasets within the cerebellum and hippocampus (Rodriques et al., 2019). It was also used to characterize spatial gene expression patterns in the Purkinje layer of mouse cerebellum. The cerebellum is marked by parasagittal bands of gene expression in the Purkinje layer that relate with heterogeneity in Purkinje cell physiology and projection targets. An antigen of Zebrin II antibody, *Aldoc*, shows similar parasagittal expression as the cerebellum but the expression variation of this gene was unknown. Using the spatial information provided by slide-seq they

identified 669 spatially non-random genes in the Purkinje layer (Rodriques et al., 2019). Of the 669, approximately 126 genes appeared either correlated or anticorrelated with the Zebrin pattern. *Aldoc* was used as a Zebrin II (+) marker and *Plcb4* as a Zebrin II (-) marker. This study indicates that slide-seq is a scalable method for obtaining spatially resolved gene expression data at a near single-cell resolution (Rodriques et al., 2019).

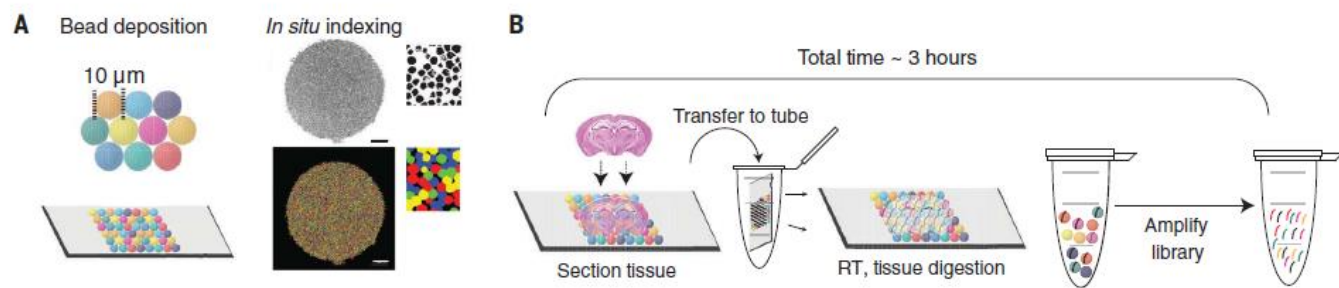


Figure 1. 12: Illustration of capturing RNA from a tissue using slide-seq.

A) Indicates the DNA barcoded beads on a glass slide forming the monolayer ('puck'). It also shows the sequenced barcodes (in black) using SoLiD sequencing. B) schematic diagram of the sample preparation procedure of slide-seq. Adapted from (Rodriques et al., 2019)

In 2020, deterministic barcoding in tissue for spatial omics sequencing (DBiT-seq) was developed, which allows for mapping mRNAs and proteins together in a formaldehyde-fixed tissues through NGS (Y. Liu et al., 2020). It does this by using polydimethylsiloxane (PDMS) microfluidic chip containing 50 parallel microchannels that delivers DNA barcodes to the surface of the tissue. This is followed by an *in situ* reverse transcription to convert the bound cDNAs into barcoded cDNAs. The tissue slide is then imaged to visualize individual pixels, each representing a unique combination of barcodes. The cDNAs are extracted, templated, and amplified for sequencing. By matching the spatial barcodes to the corresponding cDNA reads using unique molecular identifiers (UMIs), a spatial map of mRNA or protein expression is reconstructed. This map could be correlated with the tissue image to identify the spatial location of individual pixels and their associated tissue morphology. While it can provide high-resolution spatial mapping, it does not directly resolve individual cells.

Using DBiT-seq with a 25mm pixel size, Liu et al., (2020), analyzed the brain region of an E10 mouse embryo (Liu et al., 2020). Comparing results to a previous experiment with a 50mm pixel size, they observed finer structures in the pan-mRNA and pan-protein UMI count maps which correlated with the tissue morphology. They were able to identify 12 proteins out of

22, with distinct expression patterns. *CD63* showed extensive expression except in a portion of the forebrain, while *PECA*, a marker for endothelial cells, was clearly detected in brain microvasculature, which was not easily distinguishable in tissue histology. The study demonstrated the utility of high-quality spatial protein mapping data in guiding genome-wide spatial gene expression analysis.

1.7.4 10X Visium

In 2018 the ST technology was further developed by 10X Genomics under the name Visium (Fig. 1.9). Visium has improved both the resolution (from 100 μm to 50 μm) and the processing time of the original ST approach (Asp et al., 2020; Roth et al., 2020). Visium uses a spatial gene expression slide which has four capture areas, allowing for the capture of four different types of tissues on one slide. The capture area is 6.5 mm with 5000 barcoded spots. Each spot contains primers with a unique barcode which allows for the mapping of individual sequencing reads back to their specific location on the tissue (Fig. 1.13). Each spot in the capture area is 55 μm in diameter this means that each spot can capture 1 to 10 cells for analysis (Fig. 1.13). Therefore, Visium does not provide single-cell resolution. Tissue sections are placed on the different capture areas, fixed, and stained using H&E or immunohistochemistry. The tissue sections are imaged prior to capturing the RNA so that the sequencing data can be mapped to its position on the stained image (Figure 1.13).

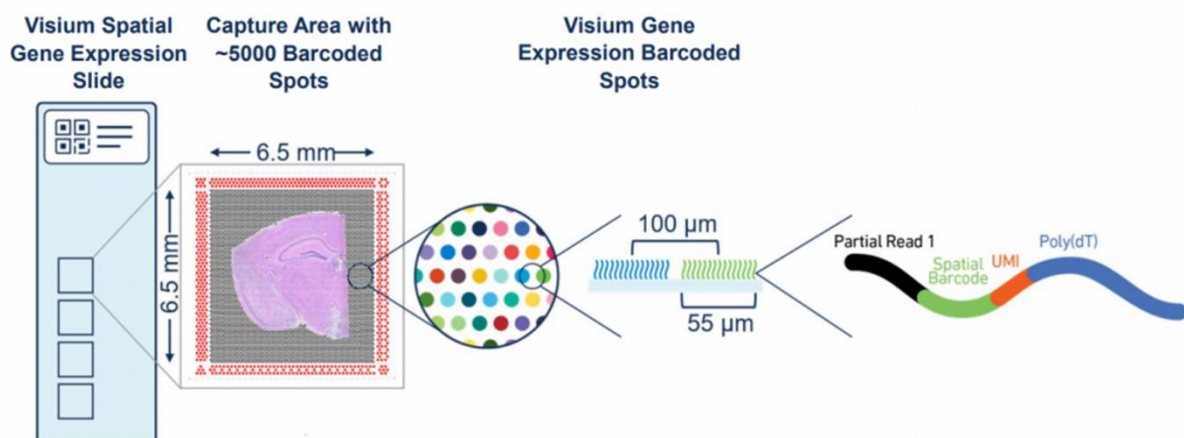


Figure 1. 13: Visium Spatial Gene Expression slide.

The slide contains capture areas containing barcoded spots of 55 μm . The capture molecules found in the spots are made up of Partial Read 1, spatial barcode, UMI and poly(dT) primers. Adapted from 10X Genomics.

1.7.5 Previous studies that used 10X Visium.

10X Visium is a fairly new technology and has been used to examine the gene expression in different tissues (Fawcner-Corbett et al., 2021; Hasel et al., 2021; Hou et al., 2021; Mantri et al., 2021). A study conducted by Maynard et al., is one of the first published articles that used Visium on the human brain. To get a better understanding of the spatial topography of gene expression in the adult human postmortem dorsolateral prefrontal cortex (DLPFC) they performed spatial transcriptomics on tissue from three adults. They used three strategies to perform differential expression analyses which used the layer-enriched expression profiles generated above the linear mixed-effects modeling. The first strategy called analysis of variance model (ANOVA) revealed the differences in gene expression between all the layers of the DLPFC beyond broad white matter (WM) verses gray matter comparisons. This strategy showed extensive differential expression across the laminar organization of the DLPFC, with 10,633 (47.6%) differential expression genes (DEGs) across the six gray matter layers plus WM and 8,581 (38.4%) DEGs across the six gray matter layers excluding WM. The second strategy called enrichment tested the difference in the gene expression in one layer versus all the other layers. The enrichment model revealed that the largest expression differences were between WM and the neocortical layers, with 9,124 DEGs and the smallest differences were between layer 3 and all the other layers with 183 DEGs. The third strategy was a pairwise analysis that tested genes that were differentially expressed between each pair of layers. This showed DEGs ranging from 8,500 for WM versus layer 3 (L3) to 292 for layer 4(L4) versus layer 5(L5). These three strategies revealed the differences in the gene expression between the different layers of the human adult DLPFC. With these 10X Visium analyses they were able to replicate previously identified layer enriched-marker genes e.g. *FABP7* which is a known layer 1 marker. The study also identified new layer enriched genes in the human cortex e.g., *TRABD2A* which they found to be enriched in layer 5.

10X Visium is an exciting and attractive technology because it has several advantages compared to other spatial transcriptomics technologies: it allows coverage across a large region of brain tissue, it does not require a selection of gene targets of interest therefore it is unbiased (Maynard et al., 2021) and it can provide total gene expression in a fresh-frozen tissue sample as there is no need for tissue dissociation. With such advantages it makes it

easier to analyze the spatial gene expression data in a single experiment; it allows for the interrogation of over 18 000 genes in a single section (Maynard et al., 2021).

A study conducted by Joglekar et al., used a modified version of the Visium protocol to validate differential isoform usage in different regions of the mouse brain. The aim of their study was to test and define cell-type-specific contributions to brain region-specific splicing and to provide a spatial exon expression map. They investigated differential isoform expression between mouse brain regions (hippocampus and prefrontal cortex) in 45 cell types at postnatal day 7 using single cell long read sequencing. In their single cell data, they showed that synaptic genes associated with vesicle transport show splicing differences during the development of the hippocampal neurons. For example, *Snap25* has two mutually exclusive isoforms: *Snap25-a* and *Snap25-b*. These *Snap25* isoforms are differentially expressed between neuronal subtypes. They showed that excitatory neurons in the CA3 region of the hippocampus have a higher proportion of *Snap25-b* transcripts than those located in the CA1 and subiculum regions of the hippocampus. Additionally, granule neuroblast clusters (GranuleNBs) showed a higher abundance of *Snap25-a* whereas mature excitatory neurons switch to *Snap25-b*. They validated these results using Visium which showed that the switch from *Snap25-a* to *Snap25-b* occurs in a posterior to anterior gradient across the hippocampus in spatially mapped transcripts (Joglekar et al., 2021).

Hasel et al., conducted a study to investigate the transcriptome of the mouse cortical astrocytes after an acute inflammatory stimulus caused by an infection from bacterial cell wall endotoxin lipopolysaccharides (LPS) (Hasel et al., 2021). They used scRNA-seq of mouse brain tissue to investigate the heterogeneity of astrocyte subtypes and their response to inflammation. Their analysis identified 10 distinct astrocyte sub-clusters. In addition to scRNA-seq, they also used the Visium Spatial Gene Expression platform to analyze the spatial distribution of astrocyte subtypes and their response to inflammation in different regions of the brain, focusing on the cortex and underlying white matter tracts. They were able to identify astrocyte-specific genes such as *Gfap*, *S100b*, or *Aldh1l1*, as well as gene "modules" from scRNA-seq experiments to approximate the location of astrocyte clusters in the brain sections. They found that different astrocyte clusters occupied distinct anatomical locations in the brain, with some located in the subpial area, while others were present in cortical gray matter or white matter tracts. For example, cluster 4 and cluster 6, both of which contained

specific markers for the deep layers of the brain (*Id3* and *Cd9*, respectively), were in the same part of the white matter tracts that contain *GFAP*-positive cells. In contrast, cluster 3, which had genes known to be markers for the grey matter of the brain (such as *Igfbp2*), was found in the cortical plate of the grey matter. They revealed that the cluster 4 genes were more highly activated in the white matter region of samples treated with LPS when compared to samples treated with saline (a control substance) (Hasel et al., 2021).

A study conducted by Orsenigo et al., used scRNA-seq in combination with Visium to fully characterize a subclass of brain Endothelial cells (ECs) to understand a rare neurovascular disease called cerebral cavernous malformation (CCM) (Orsenigo et al., 2020). They studied this using a mouse model under both normal conditions and after the depletion of *Pdcd10* (*Ccm3*) in a mouse model of CCM. They validated the scRNAseq data using Visium. They identified diverse anatomical regions of the cerebellum such as internal granular layer, and molecular layer. They identified spots that expressed pan-endothelial and lesion-endothelial markers and they found that some co-expressed *Mki67*. With Visium they were able to confirm enhanced proliferation specifically in the ECs of the cavernomas (Orsenigo et al., 2020).

To use spatial transcriptomics to explore the spatial relationships between cell types, Kleshchevnikov et al., developed a method called cell2location (Kleshchevnikov et al., 2022). They used this tool to integrate scRNA-seq and spatial transcriptomics data to map fine-grained cell types in the mouse brain. By comparing the spatial patterns of gene expression in the Visium data to the cell types from scRNA-seq data, cell2location was able to assign each gene to a specific cell type and location within the brain tissue. They found that cell2location accurately identified the location of cell types within the brain tissue and recapitulated the known layered structure of the brain. This demonstrated that cell2location is a powerful tool for mapping the spatial distribution of cell types within complex tissues. In addition to the mouse brain, the researchers also used cell2location to analyze Visium data from other tissue types, including the lymph node. Cell2location accurately identified the location of cell types within the tissue and detect regions of colocalizing cell types (Kleshchevnikov et al., 2022).

These studies have shown how powerful Visium technology is and the amount of information it can provide to scientists, especially when combined with scRNA-seq. Several studies have helped us to understand the complexity of the human brain. Very little information is known

about the spatial dynamics of the human brain, particularly the juvenile stage. Using Visium technology this study aims to understand the spatial dynamics of the human brain as it matures post-natally from childhood to adulthood.

1.8 Validating selected Visium gene expression findings using *in situ* hybridization chain reaction.

While Visium makes it possible to spatially resolve gene expression in a complex tissue, it is important to validate a subset of gene expression profiles using a targeted approach like *in situ* hybridization chain reaction (HCR). HCR is a histological method which allows multiplexed experiments to be performed with an experimental timeline independent of the number of target mRNAs (Choi et al., 2016). It is a nucleic acid-based amplification system which is used to detect specific RNA or DNA sequences in cells, preserved tissue section or an entire tissue (Zhuang et al., 2020). In HCR amplifiers will polymerize only in the presence of specific DNA sequences. A DNA probe fused to one or more of those specific DNA sequences (initiators) is hybridized to a target molecule. This is followed by subsequent incubation with the amplifiers leading to polymerization and the accumulation of label at the probe binding site (Fig. 1.14). One of the most important applications of HCR is to simultaneously localize customized panels of transcripts in tissue. This methodology has the potential to provide positional information for molecular characterization of heterogeneous cell populations in the brain (Kumar et al., 2021).

HCR has been used together with scRNA-seq, where scRNA-seq clusters cells into different cell types and generating landmark genes for each cell type (Liao et al., 2021). HCR analysis using landmark gene probes exposed the spatial distribution of their expression patterns, yielding expression densities of marker genes across the entire tissue (Liao et al., 2021).

In this study we will be using 10X Visium to analyze spatial gene expression profiles of human brain tissue across a range of ages and a subset of these results will be validated using HCR.

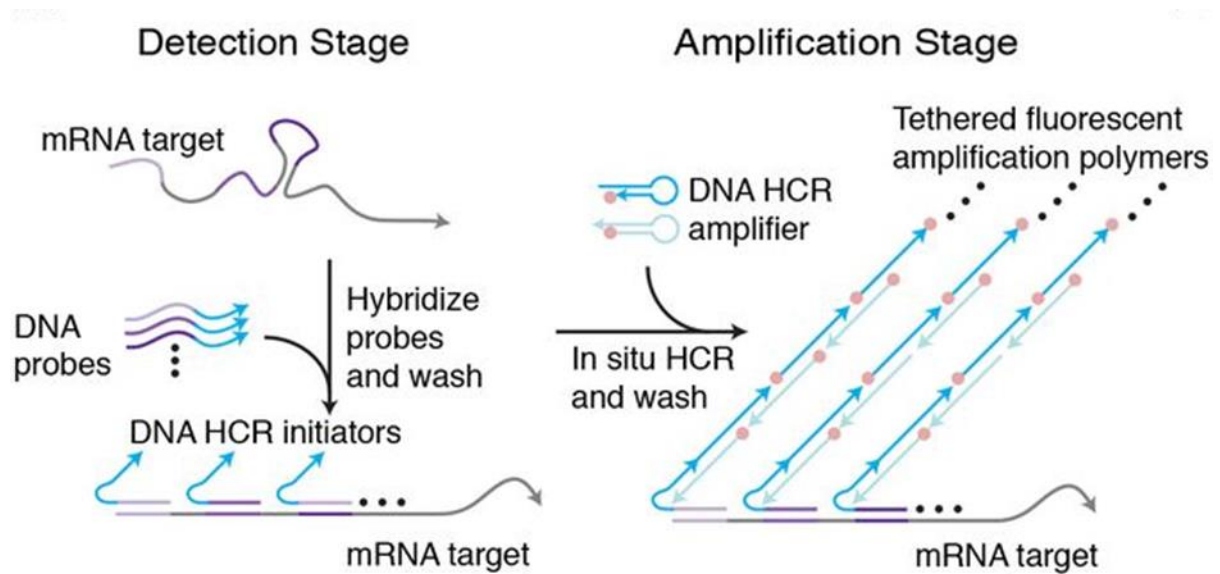


Figure 1. 14: In situ hybridization chain reaction process.

HCR involves two stages: detection stage which involves Detection stage: The probe sets carrying DNA HCR initiators hybridize to mRNA targets. The unused probes are washed from sample. Amplification stage: metastable DNA hairpins penetrate the sample. The specifically bound probe pairs trigger self-assembly of a tethered fluorescent amplification polymer which amplifies the signal. The unused hairpins are washed from sample. Adapted from (Shah et al., 2016a)

1.10 Study Aim

This study aims to contribute to the cell atlas of the maturing human brain by analyzing spatial gene expression profiles in bio-banked ante-mortem temporal cortex samples over the course of brain maturation. These analyses will reveal how gene expression varies with cell localization in the cerebral cortex as the brain matures.

1.11 Objectives

- 1) To generate spatial transcriptomics datasets using the 10x Genomics Visium platform.
- 2) To compare spatial transcriptomic datasets from juvenile and adult tissues and determine how spatial gene expression patterns and cell locations change over the course of brain maturation.
- 3) To optimize the *in-situ* Hybridization chain reaction protocol (HCR) on human brain tissue.
- 4) To confirm selected gene expression patterns identified in the Visium datasets using HCR.

CHAPTER 2: METHODOLOGY

2.1 Ethical Approval

Patient consent was required for this study. This study was conducted with the approval from the University of Cape Town Human Research Ethics Committee (Protocol No: 016-2018, sub-study 146/2022). The tissue used in this study was obtained from Mediclinic Constantiaberg and Red Cross Children's War Memorial hospitals on the condition that no personal identifiers of the patients may be collected or kept (Table 2.1). The tissue was obtained during a surgical treatment of epilepsy which involves the removal of the region of the brain where the seizures originate. The tissue obtained in this study was located outside the epileptic focus, which is likely to be non-epileptic.

Table 2.1: Patient samples used in this study: the gender, age, brief diagnosis and the types of experiment conducted for each sample.

Sample	Sex	Age	Diagnosis	Experiment	Visium Batch
P0013	Female	15	Left temporal lesion and epilepsy	Tissue optimization, snRNAseq Visium, HCR	First Batch (B1)
P0015	Male	31	Right hippocampal lesion and temporal lobe epilepsy	snRNAseq, Visium, HCR	First Batch (B1)
P0021	Male	4	Right temporal lobe cystic lesion and epilepsy	snRNAseq, Visium, HCR	Second Batch (B2)
P0029	Male	15	Left temporal lobe epilepsy	snRNAseq, Visium, HCR	Second Batch (B2)

2.2 Sample preparation

Samples were transported to the lab in ice-cold artificial cerebral spinal fluid (aCSF)(110 Mm choline chloride, 26 Mm NaHCO₃, 10 Mm D-glucose, 11.6 Mm sodium ascorbate, 7 Mm MgCl₂, 3.1Mm sodium pyruvate, 2.5 Mm KCl, 1.25Mm NaH₂PO₄, and 0.5 Mm CaCl₂) (Mohan et al.,

2015). Upon arrival at the lab, the tissue samples were dissected into small pieces while still immersed in aCSF. They were placed in plastic moulds and immediately embedded in OCT (Leica) and snap frozen to prevent RNA degradation by dipping the mould in liquid nitrogen or in a bath of isopentane that was cooled in liquid nitrogen. Some of the samples were flash frozen in liquid nitrogen. All samples were stored at -80°C until use. (Fig. 2.1)

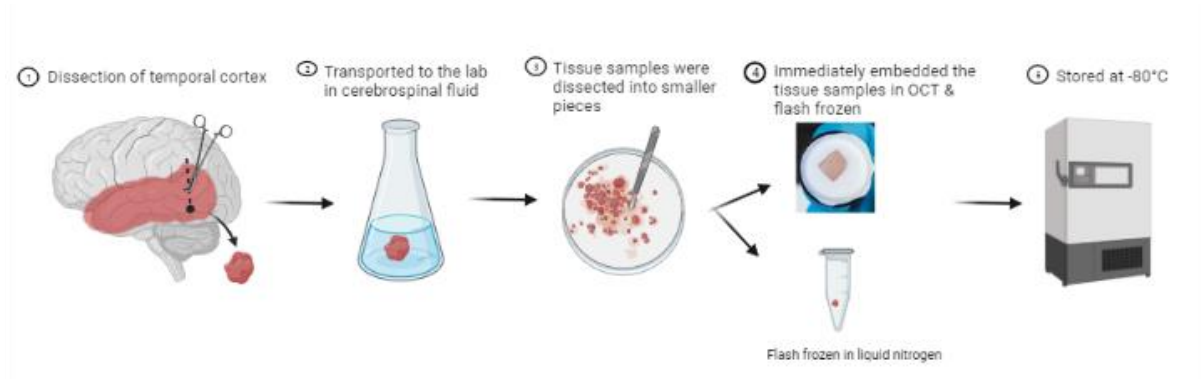


Figure 2. 1: Schematic showing the process of sample preparation.

(1) Tissue was dissected from the temporal cortex. (2) The tissue was transported to the lab in aCSF. (3) While still in aCSF the tissue was dissected into smaller pieces. (4) The smaller pieces were embedded in OCT and frozen or flash frozen in liquid nitrogen. (5) The samples were stored at -80°C. (Made in biorender)

2.3 Spatial Gene Expression Visium Experiment

The Visium Experiment consist of two steps in which the tissue is collected onto specific slides:

1) tissue optimisation step to assess the best permeabilization time for the tissue 2) the spatial gene expression step where the optimized permeabilization time is applied to make the spatial gene expression libraries.

2.3.1 RNA Quality Assessment (RNA Extraction)

RNA was extracted from a flash frozen 5-year-old tissue sample to use as a positive control in the Tissue Optimisation step. The RNAqueous Micro Kit Protocol was used. The tissue sample was thoroughly disrupted in 100 µl lysis solution using a mechanical homogenizer. 50 µl of 100% ethanol was added and vortexed thoroughly. The lysate/ethanol mixture was passed through the Micro Filter Cartridge Assembly and centrifuged for 30 sec at maximum speed until all the mixture had passed through the filter. Added 180 µl Wash Solution 1 to the filter and centrifuged for 30 sec until the solution has passed through the filter. After centrifugation, 180 µl of Wash Solution 2/3 was added to the micro filter cartridge and

centrifuged for 30 sec until the solution had passed through the filter. This was repeated with a second 180 µl of Wash Solution 2/3. The flow-through was discarded and the filter was centrifuged for 1 min at maximum speed to remove residual fluid and dry the filter. The RNA was eluted into a Micro Elution Tube by adding a pre-heated to 75°C, 10 µl Elution Solution to the centre of the filter. This was incubated at room temperature for 1 min followed by centrifuging for 30 sec to elute the RNA from the filter. This final step was repeated with a second 10 µl aliquot of the preheated Elution Solution and eluted in the same Micro Elution Tube.

2.3.1.1 DNase Treatment and DNase Inactivation

After Elution of the RNA, 1/10 volume of the 10X DNase I Buffer and 1 µl of DNase I was added to the RNA and mixed gently. This mixture was incubated for 20 min at 37°C. After incubation the DNase Inactivation Reagent was vortexed vigorously and 1/10 volume of it was added and mixed well. This was incubated at room temperature for 2 min. Followed by centrifuging for 1.5 min at maximum speed to pellet the DNase Inactivation Reagent. The RNA was transferred to a fresh RNase-free tube. The Nanodrop was used to check the RNA concentration prior to RNA Integrity Number (RIN) assessment. After Nanodrop, 2.5 µl was sent for quality check using an Agilent bioanalyser RNA 6000 Nano at the Central Analytical Facilities (Stellenbosch University). A bioanalyser is a device that utilizes microfluidics technology to perform automated and consistent electrophoretic separations for bio-analysis. It achieves this by separating small quantities of RNA samples within microfabricated chip channels based on their molecular weight, followed by detection using laser-induced fluorescence. The output is presented as an electropherogram, which represents the measured fluorescence and corresponds to the quantity of RNA of a specific size (Schroeder et al., 2006). By analysing the electropherogram, the RIN software algorithm can assess the RNA integrity. It considers the 18S to 28S ribosomal band ratio and the baseline signal between them, which change as RNA degradation occurs. The RIN software algorithm analyses the electropherogram to assign a RIN to each sample. This number, ranging from 1 (most degraded) to 10 (most intact), provides an objective measure of the RNA sample's quality and integrity. The remaining RNA was stored at -80°C until tissue permeabilization.

2.3.2 Tissue Sectioning

Both OCT embedded brain tissue samples and the Visium slides (Tissue Optimization slide and Spatial Gene Expression slides) were equilibrated to the cryostat temperature for 30 min before sectioning and were cut at a thickness of 10 μm using a cryostat (Leica CM1860/CM1950) (Fig. 2.2). The tissue samples were scored using a pre-chilled razor blade to fit in the Visium slide capture areas as described in the 10X Genomics protocol guide CG000240 Rev D. For the initial Tissue Optimisation step, the 15-year-old sample (P0013) was sectioned onto the tissue optimization slides (Table 2.1). For the Visium Spatial Gene Expression experiment, two batches of sample collection were performed. For the 1st Visium batch two replicate sections of the 15-year-old (10 μm apart) and two replicate sections of 31-year-old (40 μm apart) were collected onto the Spatial Gene Expression slide. For the 2nd Visium batch two replicate sections of the 2nd 15-year-old (10 μm apart) and two replicates of the 4-year-old were used (adjacent sections) (Table 2.1). The spatial Gene Expression slides with tissue sections were stored individually in a sealed container at -80°.

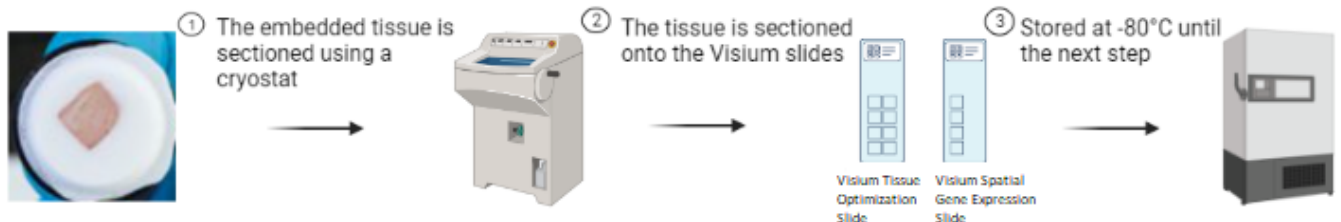


Figure 2. 2: Schematic representation of the steps involved in tissue sectioning.

(1) The OCT embedded sample is taken out of the -80°C freezer and sectioned using a cryostat. **(2)** The tissue is collected onto the Visium slides. **(3)** The Visium slide is stored in the -80°C freezer. (Made in biorender)

The 10x Genomics Visium Spatial gene expression system was used to obtain the spatial gene expression information of the paediatric and adult temporal tissue samples. Visium measures the total mRNA in a tissue sample and maps it back to its location as described (see Section 1.7.4 10X Visium). All reagents were provided by 10X Genomics unless stated otherwise.

2.3.3 Hematoxylin & Eosin H & E

To record the morphology of the tissue samples Hematoxylin and Eosin (H&E) staining was performed on both the Tissue Optimization slide and Spatial Gene Expression slides as

detailed in the 10x Genomics demonstrated protocol guide CG000160 Rev B (Fig. 2.3). Slides with sections were transported from the -80°C freezer on dry ice and were immediately incubated at 37°C for 1 min on a heated Thermocycler C1000 Touch Thermal Cycler (Bio-Rad Laboratories). The sections were fixed by incubating in a chilled methanol (Kimix Chemical and Lab Supplies) for 30 min at -20°C, followed by adding 500 µl of isopropanol to the sections at room temperature for 1 min. The sections were air dried for 5-7 min and stained with 1000 µl hematoxylin (Sigma-Aldrich) for 7 min. Slides were dipped in three UltraPure distilled water (Invitrogen) beakers for 5x, 15x and 15x respectively. This was followed by adding 1000 µl bluing buffer (Dako) for 2 min. Slides were immersed in UltraPure water beaker 2 for 15x. Eosin (Sigma-Aldrich) mix was added onto slide for 1 min. After immersing slides in UltraPure water beaker 3 for 15x, the slides were incubated at 37°C for 5 min.

Prior to imaging the tissue optimization slide and the 2nd Visium batch slide, the slide was mounted with a mounting medium containing 85% glycerol (Sigma-Aldrich), 5 mM RNase Inhibitor (Ambion) and Nuclease free water and covered with a glass coverslip. The 1st Visium batch slide was imaged without a coverslip. Brightfield images of the sections were captured using an EVOS M5000 microscope (Thermo Fisher Scientific) at 20x magnification. The images were tile scanned manually by taking overlapping images of the section including the fiducial frame. After imaging the images were stitched together using Image Composite Editor-2.0.3 (Microsoft). For the 2nd Visium experiment, the coverslip was removed by immersing the slide in 3X Saline Sodium Citrate (SSC) Buffer (Invitrogen) until the coverslip separated from the slide.

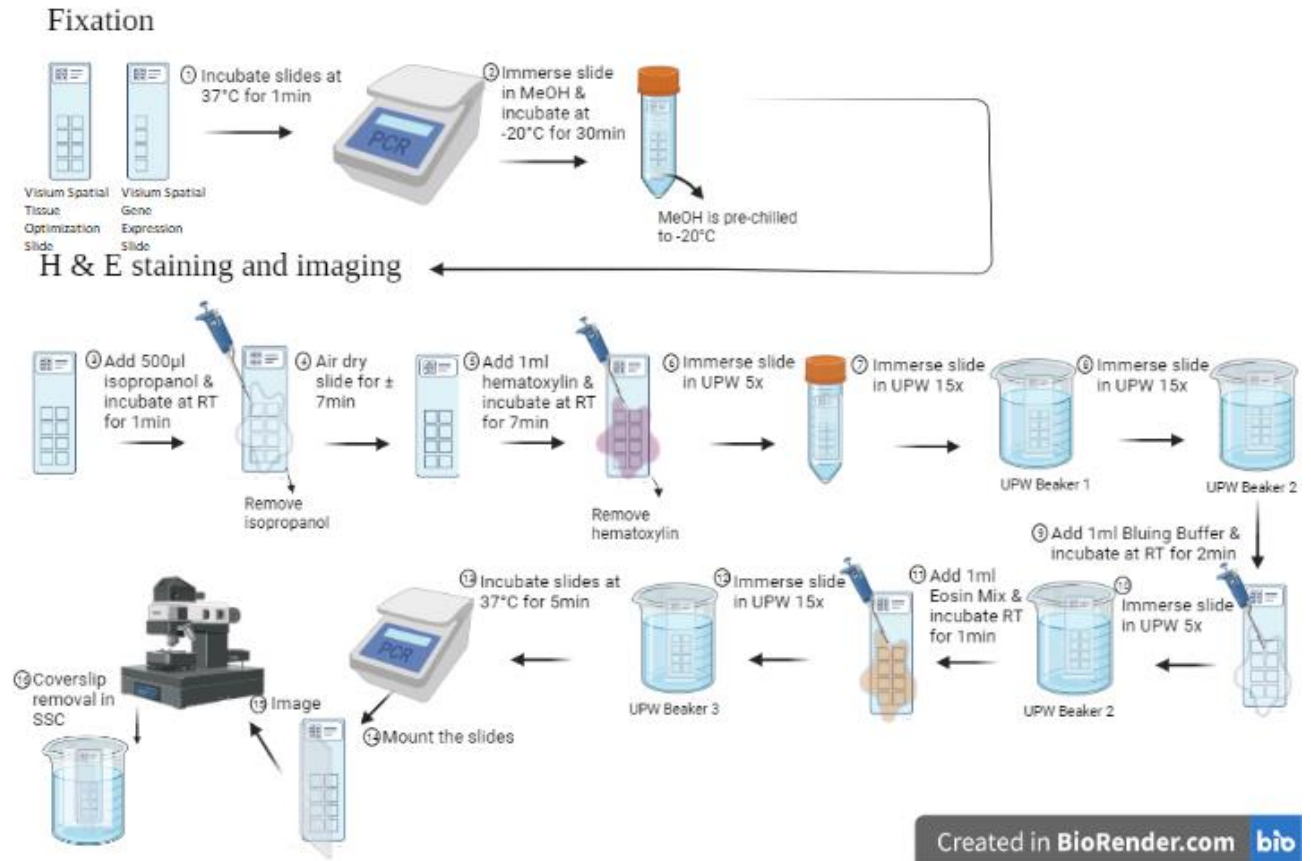


Figure 2. 3: Illustration of the H&E staining procedure used.

(1,2) The Visium slide is fixed in pre-chilled methanol. **(3,4)** The slide was incubated in isopropanol at room temperature and air dried. **(5)** Incubated slide in haematoxylin at room temperature. **(6,7,8)** After incubation the slide was washed in ultra-pure water. **(9,10)** A bluing buffer was added to the slide and incubated at room temperature for 2min. The bluing buffer was removed and rinsed in ultra-pure water. **(11)** Incubated slide in eosin mix at room temperature. **(12)** Rinsed the slide in ultra-pure water. **(13)** Dried slide by Incubating at 37°C. **(14,15)** Mounted the slide and imaged. **(16)** Coverslip removal in SSC. (Made in biorender)

2.3.4 Tissue Permeabilization Optimization

This step is required for optimization of tissue permeabilization conditions prior to conducting Visium library generation (Fig. 2.4). This step was performed on 15-year-old tissue (P0013) (Table2.1), which was collected onto the Visium Spatial Tissue Optimization Slide as detailed in Section 2.3.2. The 10x Genomics Visium Spatial Tissue Optimization Reagent Kit (PN-1000192 & PN-1000191) was used according to the instructions in the user guide CG000238 Rev C.

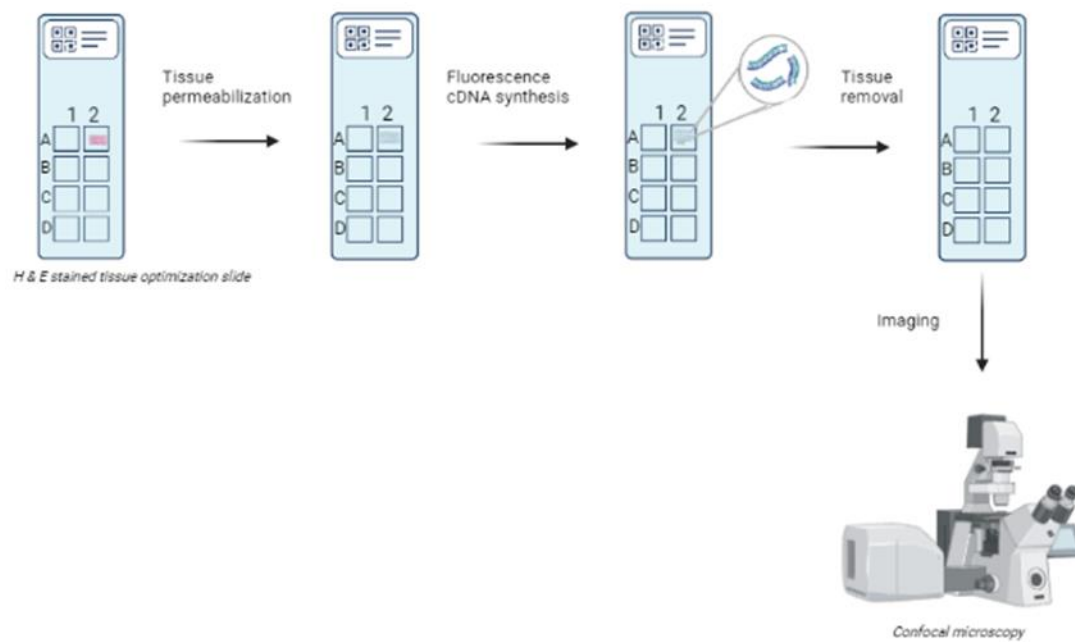


Figure 2. 4: Schematic showing the steps involved in the Visium tissue optimization step.

Step 1: The tissue is permeabilized. **Step 2:** Fluorescence cDNA synthesis. **Step 3:** Tissue removal. **Step 4:** Imaging using the confocal microscope. (Made in biorender)

After H&E staining and imaging as detailed in section 2.3.2 the slide was placed in the provided slide cassette (Fig. 2.5), creating a separate well over each section. 1 $\mu\text{g}/\mu\text{l}$ RNA from the 5-year-old flash frozen tissue piece (see Section 2.3) was added to the positive control well. Permeabilization enzyme was added to the each of the wells except for negative control and positive control wells, which did not contain tissue. The slide seal was applied to the slide cassette and placed on a pre-heated thermocycler adaptor in the thermal cyclor, which was set to 37°C for the duration of the permeabilization time course. The tissue in each well was permeabilized for a specific period (30 min, 24 min, 18 min, 18 min, 12 min, 6 min and 3 min) (Fig. 2.5).



Figure 2. 5: An image of the 10X Genomics slide cassette used in the Visium experiment.

The slide cassette indicates the layout of permeabilization time course. (adapted from Visium tissue optimization guide, CG000238 Rev E)

After the time course was complete, the slide seal and permeabilization enzyme was removed and the wells were washed with 100 μ l 0.1X SSC, except the positive control. The 0.1X SSC was removed and 50 μ l Fluorescent RT Master Mix was added to each well (Table 2.2). The slide seal was applied to the slide cassette and placed on the thermocycler adaptor on the pre-heated thermal cycler to start fluorescent cDNA Synthesis (Table 2.3).

Table 2.2: Fluorescent RT Master Mix: Reagents volumes used to make the master mix for fluorescent cDNA synthesis.

Reagents	Volume (μ l)
Nuclease Free Water	221.8
RT Reagent C	110.0
Template switch oligo	30.8
Reducing Agent B	8.8
RT Enzyme D	68.6
Total	440

Table 2.3: Fluorescent cDNA synthesis thermal cycler conditions: The steps involved in fluorescent cDNA synthesis, showing the temperature in each step and the duration of each step.

Lid Temperature		Run Time
53°C		45min
Step	Temperature	Time
Pre-equilibrate	53°C	Hold
cDNA Synthesis	53°C	45min
Hold	4°C	

After fluorescent cDNA Synthesis, the slide seal was removed, and the fluorescence RT Master Mix was removed from the wells and 100 µl 0.1X SSC added to each well. The 0.1X SSC was removed and 70 µl Tissue Removal Mix (Table 2.4) was added to each well. The slide seal was applied to the slide cassette and placed on the thermocycler adaptor on the pre-heated thermal cycler to start tissue removal (Table 2.5).

Table 2.4: Tissue Removal Mix: Reagent volumes used to make the mix for the tissue removal step.

Reagents	Volume (µl)
Tissue Removal Buffer	539
Tissue Removal Enzyme	77
Total	616

Table 2.5: Tissue removal thermal cycler conditions: The steps involved in tissue removal, showing the temperature in each step and the duration of each step.

Lid Temperature		Run Time
56°C		60min
Step	Temperature	Time
Pre-equilibrate	56°C	Hold
Tissue Removal	56°C	60min
Hold	22°C	

After tissue removal incubation period, the tissue removal mix was removed from the wells and the slide was removed from the slide cassette. The slide was immersed 15x in pre-warmed 2X SSC- 0.1% SDS at 50°C, followed by immersed 15x immersions in both 0.2X SSC

and 0.1X SSC buffers at room temperature. The slide was placed in a 50 ml centrifuge tube and centrifuged for 30sec at 250 RCF (Eppendorf Centrifuge 5702 R).

2.3.4.1 Imaging

A coverslip was mounted on the slide for imaging. To determine the optimum permeabilization time, fluorescence imaging was performed using the LSM 880 Airyscan confocal microscope (Carl Zeiss, ZEN SP 2 software) using the 20X objective (Zeiss) following the Visium Spatial Gene Expression Imaging Guidelines CG000241 Rev D. Prior to fluorescence imaging, the Visium imaging test slide was used to verify imaging settings needed to obtain a clear scan of the entire slide (Fig. 2.6). The fluorescent spots in all the test slide areas were visible, decreasing in brightness from left to right as shown in Fig. 2.6. These settings were saved and used for scanning the tissue optimization slide. The fluorescence output across all the frames was analysed using the Zen Black software version 3.3, to be able to identify the frame with the brightest fluorescence signal. The permeabilization time of the brightest frame was used for subsequent steps.

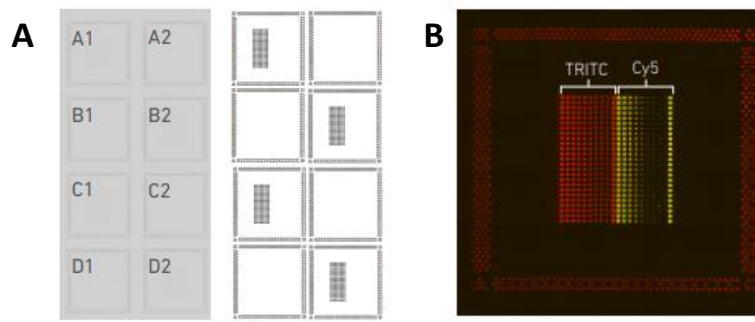


Figure 2. 6: An image showing the Visium test imaging slide.

(A) Indicates where the fluorescent spots are located on the slide. (B) Shows the fiducial frame and the fluorescent spot decreasing from left to right which is seen with fluorescence imaging. Adapted from Visium Spatial Gene Expression Imaging Guidelines CG000241 Rev D.

2.4. The Visium library preparation

2.4.1 Tissue permeabilization and Reverse Transcription

The tissue samples (Table 2.1) were sectioned onto Visium slides (Fig. 2.7), stained with H&E and imaged (see Sections 2.3.3. For the 2nd Visium experiment, the coverslip on the slide was removed as described in section 2.3.3. For both Visium experiments, the slide was placed in the slide cassette and washed with wash buffer. The wash buffer was removed and 70 µl permeabilization enzyme was added to the wells. The tissue was permeabilized at 37°C as described in section 2.3.4 using the optimised permeabilization time of 12 min. After permeabilization the permeabilization enzyme was removed and 0.1X SSC was added. This was followed by reverse transcription, by first removing the SSC and adding 75µl RT Master Mix (Table 2.2). The slide cassette was sealed with a slide seal and placed onto the thermocycler adaptor in the thermal cycler for reverse transcription (Table 2.3)

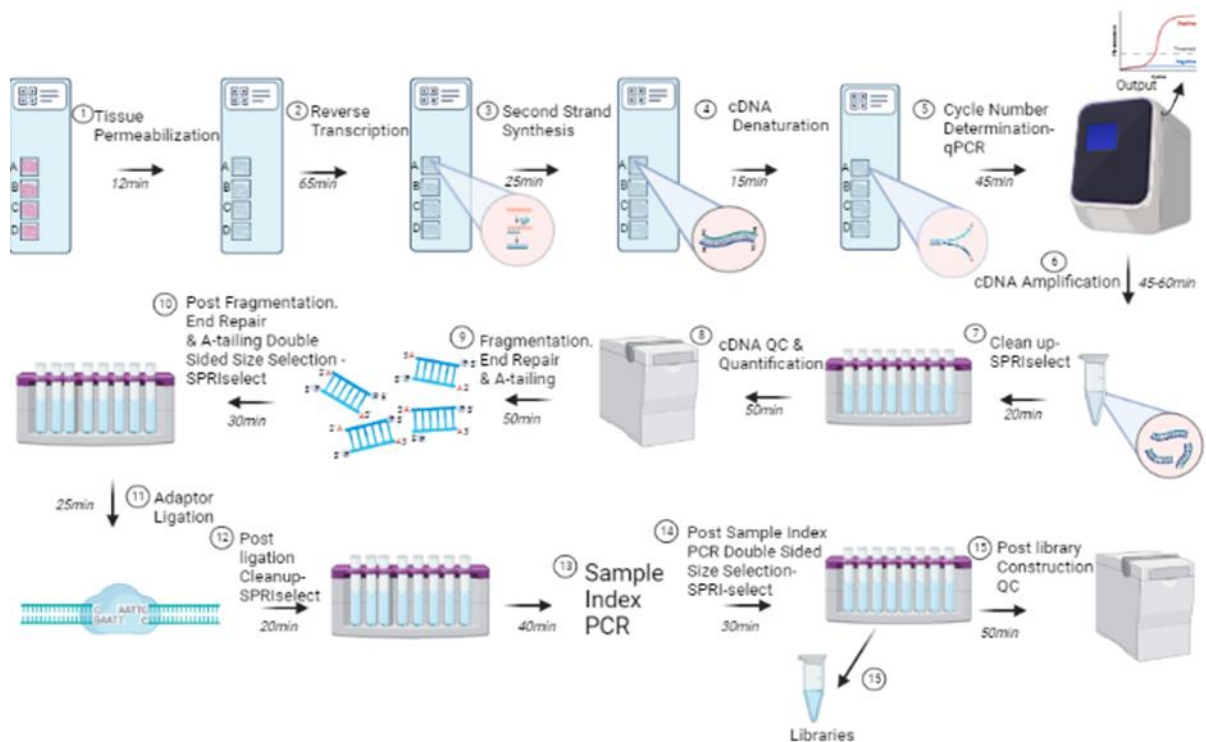


Figure 2. 7: The diagram showing the Visium protocol steps & timing.

(1,2) Tissue permeabilization using the optimized time and reverse transcription. **(3,4)** Second Strand Synthesis and cDNA Denaturation. **(5-8)** cDNA Amplification & QC. The samples are analysed used qPCR to determine the number of cycles used in the amplification step of the cDNA via PCR. **(9-15)** Spatial Gene Expression Library Construction. This involves enzymatic fragmentation and size selection of the cDNA. End Repair, A-tailing, ligation and PCR. (Made in biorender)

2.4.2 Second Strand Synthesis

After reverse transcription (Fig. 2.7) the slide cassette was removed from the thermal cycler and the RT Master Mix was removed from the wells. 0.08M KOH was added to the wells and incubated for 5 min at room temperature. After incubation the KOH was removed and 100 μ l buffer EB (Qiagen) was added. The EB was removed and 75 μ l Second Strand Mix (Table 2.6) was added to each well. The slide seal was applied to the slide cassette and transferred to the preheated thermocycler adaptor for second strand synthesis (Table 2.7).

Table 2.6: Second Strand Mix. The reagent volumes used to make the mix for second strand synthesis.

Reagents	Volume (μ l)
Second Strand Reagent	305.8
Second Strand Primer	17.6
Second Strand Enzyme	6.6
Total	330.0

Table 2.7: Second Strand synthesis thermal cycler conditions: The steps involved in second strand synthesis, showing the temperature and duration in each step.

Lid Temperature		Run Time
65°C		15min
Step	Temperature	
Pre-equilibrate	65°C	Hold
Second Strand Synthesis	65°C	15min
Hold	4°C	Hold

After second strand synthesis the second strand mix was removed and 100 μ l buffer EB was added to each well. The buffer was removed, 35 μ l 0.08M KOH was added to each well and incubated at room temperature for 10min. After incubation 35 μ l of the sample from each well was transferred to a separate sample tube containing 5 μ l 1M Tris (pH 7.0). This mixture was vortexed, centrifuged and placed on ice.

2.4.3 cDNA Amplification

qPCR was performed to determine the number of cycles to use for cDNA amplification (Fig. 2.7). 9 μ l of the qPCR mix (Table 2.8) was added to 1 μ l of each sample from section 2.4.2. The negative control contained 1 μ l nuclease free water instead of sample. qPCR (Table 2.9) was

performed on the LightCycler qPCR machine (Roche). The C_q value of each sample was recorded after qPCR, with the threshold set along the exponential phase of the amplification plot at 25% of the peak fluorescence value. This C_q value was used to determine the cycles for cDNA amplification, which was 20 cycles. 65 µl cDNA amplification mix (Table 2.10) was added to the remaining 34 µl sample, and cDNA amplification performed in a thermocycler (Table 2.11).

Table 2.8: qPCR Mix. The reagent volumes used to make the mix for qPCR.

qPCR Mix reagents	Volume (µl)
Nuclease-free Water	20.4
KAPA SYBR FAST qPCR Master Mix	27.5
cDNA Primers	1.7
Total	49.6

Table 2.9: qPCR conditions: The steps involved cDNA amplification, showing the temperature and duration in each step.

Lid Temperature	Reaction Volume	Run Time
-	10µl	35min
Step	Temperature	Time
1	98°C	3min
2	98°C	5sec
3	63°C	30sec
3	For 25 cycles	-

Table 2.10: Amp Mix. The reagent volumes used to make the mix for cDNA amplification.

Reagents	Volume (µl)
Amp Mix	220
cDNA primers	66
Total	286

Table 2.11: cDNA amplification thermal cycler conditions: The steps involved cDNA amplification, showing the temperature and duration in each step.

Lid Temperature	Reaction Volume	Run Time
105°C	100µl	45min
Step	Temperature	Time
1	98°C	3min
2	98°C	15sec
3	63°C	20sec
4	72°C	1min
5	20 cycles	
6	72°C	1min
7	4°C	Hold

cDNA amplification was followed by sample clean-up to remove amplification reagents and primers using the SPRIselect reagent (Beckman Coulter). SPRIselect is a type of chemistry that utilizes SPRI (Solid Phase Reversible Immobilization) technology to make the process of selecting nucleic acid sizes faster and easier for preparing fragment libraries for Next Generation sequencing (Stortchevoi et al., 2020). During this process, it is necessary to selectively choose fragments of a specific size to achieve a consistent distribution around an average size. By employing SPRIselect, the size distribution can be customized to match the specific requirements of the sequencing application and platform being used (Stortchevoi et al., 2020). Additionally, this process is adaptable to different workflow capacities, ranging from low to high throughput. 40.5 µl SPRIselect reagent was added to each sample tube and pipette mixed. This was incubated at room temperature for 5 min. After incubation the tubes were placed on a magnetic rack until the solution cleared. The supernatant was removed, 80% ethanol added to the bead pellet and removed after 30 sec. This was repeated for a total of two washes. The bead pellets were air dried for 2 min and 40.5 µl buffer EB added to each tube and mixed. After 2 min incubation at room temperature. The tube was added back onto the magnetic rack until the solution cleared and 40 µl of each sample was transferred into new tubes for storage at -20°C. 5 µl was sent for cDNA QC & Quantification using an Agilent TapeStation at the Central Analytical Facilities (Stellenbosch University). After receiving the cDNA QC & Quantification results, the cDNA Total Yield was calculated using the following

calculation:
$$\frac{[\text{ng} \mu\text{l}^{-1}] \times \text{Elution Volume}(\mu\text{l})}{1000(\frac{\text{ng}}{\mu\text{l}})}$$

2.4.4 Spatial Gene Expression Library Construction

After cDNA library QC, the final libraries were prepared through a process of fragmentation, end repair and A-tailing, followed by adapter ligation and addition of sample-specific indexes for sequencing (Fig. 2.7). For fragmentation end repair and A-tailing, 25µl buffer EB and 15 µl fragmentation mix (Table 2.12) was added to 10 µl of the purified cDNA sample, except for the 2nd Visium experiment 15-year-old (T2) which had a lower total yield than the rest of the samples. For this sample 20 µl of the purified cDNA was used and 15 µl EB was used. The tubes were transferred into the pre-cooled thermal cycler for fragmentation (Table 2.13).

Table 2.12: Fragmentation Mix. The reagent volumes used to make the mix for Library Construction.

Fragmentation Mix reagents	Volume (µl)
Fragmentation Buffer	22
Fragmentation Enzyme	44
Total	66

Table 2.13: Library Construction thermal cycler conditions: The steps involved in Library Construction, showing the temperature and duration in each step.

Lid Temperature	Reaction Volume	Run Time
65°C	50µl	35min
Step	Temperature	Time
Pre-cool block	4°C	Hold
Fragmentation	32°C	5min
End Repair & A-tailing	65°C	30min
Hold	4°C	Hold

The cDNA was then cleaned up using double-sided fragment selection with the SPRIselect reagent. First, 30 µl (0.6X) of SPRIselect reagent was added to each sample and pipette mixed. This was incubated at room temperature for 5min. After incubation the tubes were placed on the magnetic rack until the solution cleared. 75 µl of the supernatant was transferred to new tubes and 10 µl (0.8X) SPRIselect was added. This was incubated at room temperature for 5min and placed back on the magnetic rack until solution cleared. The beads were then ethanol washed as described above single-sided selection followed by elution in 50.5 µl buffer EB. Adaptor ligation mix (Table 2.14) was then added to 50.0 µl of the samples, mixed and incubated in a thermal cycler (Table 2.15). After adapter ligation, the cDNA was once again

cleaned-up as described above, using single-sided fragment selection with 80 µl (0.8X) of SPRIselect reagent, and eluted from the beads in 30.5 µl buffer EB.

Table 2.14: Adaptor Ligation Mix. The reagent volumes used to make the adaptor ligation mix.

Adaptor ligation Mix	Volume (µl)
Ligation buffer	88
DNA ligase	44
Adaptor Oligos	88
Total	220

Table 2.15: Adaptor Ligation thermal cycler conditions: The steps involved in adaptor ligation, showing the temperature and duration in each step.

Lid Temperature	Reaction Volume	Run Time
30°C	100µl	15min
Step	Temperature	Time
1	20°C	5min
2	4°C	Hold

2.4.5 Sample Index PCR

Unique sample indexes were chosen from the 10x Genomics Dual Index TT Set A for each sample to prevent any overlap between samples (Fig. 2.7). 50 µl Amplification mix and 20 µl of the chosen sample index was added to 30 µl sample and mixed. The samples were incubated in a thermal cycler for 14 cycles of amplification (Table 2.16). The cycle number chosen was based on the 25% carry forward cDNA yield calculated during the post cDNA amplification QC and quantification for batch 1. For consistency, the same number of cycles were used for batch 2.

Table 2.16: Sample Index PCR thermal cycler conditions: The steps involved in PCR, showing the temperature and duration in each step.

Lid Temperature	Reaction Volume	Run Time
105°C	100µl	30min
Step	Temperature	Time
1	98°C	45sec
2	98°C	20sec
3	67°C	300sec
4	72°C	20sec
5	14 cycles	
6	72°C	1min
7	4°C	Hold

Following the addition of sample indexes, the libraries were cleaned up with double-sided fragment selection using the SPRIselect reagent as described above, firstly with 0.6X of SPRIselect reagent and secondly with 0.8X SPRIselect reagent. The cleaned library was eluted from the beads in 30.5µl buffer EB. 30µl of each sample was transferred into new tubes for storage at -20°C. 5µl was sent for cDNA QC & Quantification using the Agilent TapeStation and Qubit at the Central Analytical Facilities (Stellenbosch University) to determine the final library concentration.

2.4.6 Library Sequencing

The H&E images were processed using the 10X Loupe Browser 4.0 Visium Manual Alignment Wizard analysis to determine the required sequencing depth by calculating the number of spots that were covered by the tissue section (Fig. 2.8). This was done by aligning the image to the fiducial marker locations using the lasso tool in the Loupe Browser. 10 µl of each sample (section 2.4.4.5) was sent for sequencing (150 bp paired end) on the Illumina NovaSeq platform at Novogene to obtain at least 63 Gb data per sample.

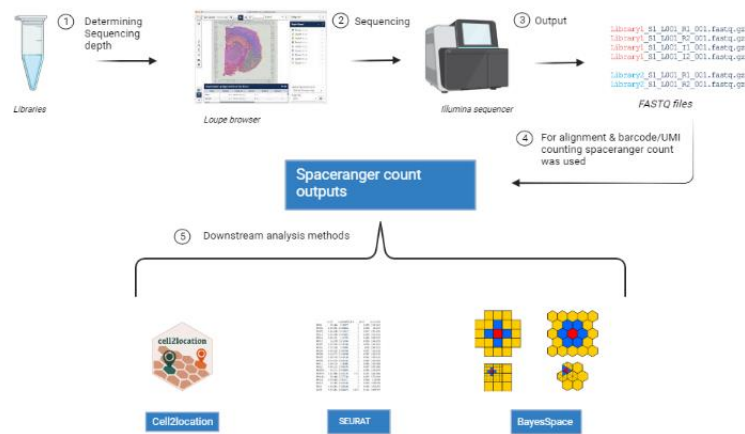


Figure 2. 8: The diagram shows an overview of the data analysis steps used in this study.

(1) Visualizing the H & E images on loupe browser to determine the sequencing depth needed for sequencing. (2,3) The libraries were sequenced using the Illumina NovaSeq platform and the outputs were FASTQ files. (4,5) Analysing the data by first aligning the data using spaceranger count and further analysing using cell2location, Seurat and BayesSpace. (Made in biorender)

2.5 Data Analysis

2.5.1 Raw data analysis

10X Genomics *Spaceranger count* (10X Space Ranger 1.3.0) was used to perform alignment, tissue detection, fiducial detection, and barcode/UMI counting using the FASTQ files obtained from Novogene as input: (`spaceranger count --id=sampleID -- transcriptome=/path/refdata/GRCh38-2020-A -- fastqs=/path/to/fastq/outs/folder -- sample=sampleID --image=/path/to/images/sampleimage.tiff --slide=Slide ID -- area=Capture area --localcores=8 --localmem=64`). The alignment file from Loupe Browser obtained as described in Section 2.4.6 and the hg38 human reference transcriptome (refdata-gex-GRCh38-2020-A) were used as additional parameters in the command. The outputs of *spaceranger count* were used for further analysis.

2.5.2 Histology image and segmentation (Vistoseg)

To determine the number of nuclei per Visium spot, the *Vistoseg* method was used. In brief, *Vistoseg* is a MATLAB pipeline that allows segmentation, analysis, and visualization of Visium data to obtain the nuclei number per spot (Tippani et al., 2021). The first step in *Vistoseg* is to perform nuclei segmentation of individual H&E images: (`VNS (fname = '/path_to_image';`

,N=number_of_colours_visually_seen_on_H&E_image)). This is followed by refining the image for accurate nuclei detection: *refineVNS* (fname = '/path_to_image'; M=index_of_the_colour_cluster_of_the_nuclei (obtained from the VNS function)). The last step is to use the *countNuclei* function which generates nuclei count csv file. The function requires these inputs: *mask* = Filename with full path for the .mat file of the final nuclei segmentation, *jsonname* = Filename with full path for the scalefactors_json.json file obtained from spaceranger *count*, *posname* = Filename with full path for the tissue_positions_list.csv file obtained from spaceranger *count*. (*countNuclei(mask,jsonname,posname)*) (see github script *Vistoseg_matlab.r*). Once *Vistoseg* was performed on all the samples, the average number of nuclei per Visium location was obtained and this was used as a hyperparameter in spatial mapping of cell types. (<http://research.libd.org/VistoSeg/step-4-gui-to-count-nuclei-in-a-Visium-spot.html>)

2.5.3 Annotation and spatial mapping of cell types (*Cell2location*)

The *Cell2location* (version 0.7a0) tool was used to spatially map the brain cell types by integrating the Visium data mRNA count matrices with existing single nucleus (sn)RNAseq datasets from the Hockman lab. *Cell2location* is a Bayesian model that estimates the cell abundance of each cell in a location by decomposing mRNA counts in Visium data using the transcriptional signatures of reference cell types (Kleshchevnikov et al., 2022). The reference snRNA-seq dataset used here consisted of several integrated temporal lobe snRNA-seq datasets that had been previously analysed and annotated by Christina Steyn in the Hockman lab using the current published human middle temporal gyrus cell atlas (Hodge et al., 2019) as a guide, resulting in 53 annotated cell types. This reference included datasets from the same patients used in the Visium experiment, as well as published datasets (Table 2.17).

To avoid mapping artifacts, mitochondrial genes were removed from the Visium datasets prior to spatial mapping. *Cell2location* was run using python v3.0.6. Reference signatures of the 54 annotated cell populations were derived using a negative binomial regression model using the default values (see github script *creating_reference_signatures.py*). Unnormalized and untransformed snRNA-seq mRNA counts were used as input in both the regression model for estimating the reference signatures. The snRNA-seq mRNA counts were filtered to 13,870 genes and 176,012 cells: *selected = filter_genes(adata_ref, cell_count_cutoff=5, cell_percentage_cutoff2=0.03, nonz_mean_cutoff=1.12)*. The *cell2location* model for

estimating the spatial abundance of cell populations was filtered to 13,858 genes that were in both the snRNA-seq and Visium data (see github script `creating_visium_anndata.py`; `creating_model.py`). The two batches of the Visium experiments were analysed together. The following cell2location parameters were used: training iterations 30, 000 cell per location, $N^{\wedge} = 7$ (estimated using Vistoseg results), Normalization (ys) alpha prior = 20. To be able to visualise the cell abundance in spatial coordinates 5 % quantile of the posterior distribution was used, which represents the value of cell abundance that the model has high confidence in.

Cell2location's Non-negative Matrix Factorization (NMF) was used to identify cellular compartments and cell types that co-locate from the cell type abundance estimates. This analysis considers the fact that multiple cell types can co-exist at the same Visium location. In this study NMF was used to understand the tissue organisation and predict cellular interactions by training NMF for a range of factors: $R=5$ to 30. $R = 15$ was selected as a balance between capturing fine-grained clusters and splitting known well-established tissue zones (see github script `cell2loc_pipeline.py`).

Cell2location's Leiden clustering was used to identify the different regions within the tissue using the estimated cell abundance values. First the KNN graph was computed using `n_neighbours = 21`. This was followed by Leiden clustering using a resolution of 0.8, yielding 20 clusters that clearly shows the different cortical layers of the 31-year-old sample. This was then applied to the rest of the samples. The clusters were visualized by a Uniform Manifold Approximation and Projection (UMAP). UMAP is a dimension reduction technique (McInnes et al., 2018). UMAP offers a number of advantages over t-SNE; it is faster in generating the projection of the data and often preserves the global structure of the data more accurately. This means that UMAP captures not only global relationships between distinct groups but also local relationships within groups of transcriptomes. By preserving both local and global relationships, UMAP enhances our understanding of gene interactions and their impact on cellular processes (Dorrity et al., 2020). Leiden clustering also provides information on the groups of cell types that are found in the same region/ cortex layer. (see github script `cell2loc_pipeline.py`)

The cell2location tutorial was used for this section (https://cell2location.readthedocs.io/en/latest/notebooks/cell2location_tutorial.html).

Table 2.17: Existing temporal lobe snRNA-seq datasets used to build an integrated reference dataset for cell2location. The table shows the samples in the dataset, the number of replicate datasets for each sample and the dataset origin.

Age	Technical replicates	Origin
Adult_20_yr_old	2	Thrupp et al. 2020 Cell Reports
Adult_24_yr_old	2	Thrupp et al. 2020 Cell Reports
Adult_26_yr_old	1	Thrupp et al. 2020 Cell Reports
Adult_31_yr_old	2	Hockman lab unpublished dataset
Adult_41_yr_old	1	Hockman lab unpublished dataset
Adult_50_yr_old	1	Thrupp et al. 2020 Cell Reports
Juvenile_15_yr_old	7	Hockman lab unpublished dataset
Juvenile_4_yr_old	1	Hockman lab unpublished dataset
Juvenile_5yr_old	2	Hockman lab unpublished dataset
Juvenile_7yr_old	2	Hockman lab unpublished dataset
Juvenile_9_yr_old	2	Hockman lab unpublished dataset

2.5.4 Seurat FindAllmarkers

The function *findallmarkers* from Seurat (version 4.0.6) (Hao et al., 2021) was used to identify differentially expressed genes for the leiden clusters that were identified by cell2location

(Section 2.5.3). To be able to do this, a Seurat object of each Visium sample was created by following the “Analysis, visualization, and integration of spatial datasets” pipeline (https://satijalab.org/seurat/articles/spatial_vignette.html). Mitochondrial genes were regressed out when creating the Seurat objects (`mt.genes <- rownames(DH3)[grep("^MT-", rownames(DH3))]`). The filtered datasets were normalized using `sctransform` function which removes technical artifacts while preserving biological variance. The Leiden cluster numbers for each spot, for each sample, were extracted and added onto the Seurat objects (see script `Update_findallmarkers.r`).

A heatmap of the top 20 markers for each cluster was plotted using `dittoHeatmap`: (`markers %>% group_by(cluster) %>% slice_max(n = 2, order_by = avg_log2FC)` (`top20 <- markers %>% group_by(cluster) %>% top_n(n = 20, wt = avg_log2FC)`)

`dittoHeatmap(Ave_exp, top20$gene, Annot.by = c("sample", "Region_cluster"), order.by = c("Region_cluster", "sample"), slot = "data", scale = "row", fontsize_row = 4.0).`

The selected marker genes of the cell types and layer expression profiles were further explored using the unpublished Hockman lab snRNA-seq datasets and publicly available datasets, including the Allen Brain Map (Bakken et al., 2021), the human protein atlas (Karlsson et al., 2021) and SpatialLIBD datasets (Maynard et al., 2021).

2.5.5 Enhancement of spatial transcriptomic data visualisation (BayesSpace)

BayesSpace was used to show the gene expression level for selected genes per spot. BayesSpace is a Bayesian model that uses information from spatial neighbourhoods to provide a high-resolution readout (subspot level) for spatial transcriptomic data and clustering (Zhao et al., 2021). The raw gene expression counts from `spaceranger count` were normalized, log transformed and, PCA was performed on the top 2000 highly variable genes (HVGs). To be able to obtain the high-resolution gene expression, the PC values were mapped back to their original log-transformed gene expression space (spot level) using the default Bayesspace regression. To do this the PCs from the original data were used as predictors in training the model for each gene, in which the results were the measured gene expression at the spot level. The trained model was then used to predict the gene expression at sub spot level using high resolution PCs. The high-resolution model was trained using default values provided by the BayesSpace tutorial

(<https://edward130603.github.io/BayesSpace/articles/BayesSpace.html>) except for the following parameters: 7 PCs, Number of clusters = 8, nrep = 100 000, burn-in = 10 000. (see github script Bayesspace.r)

2.6 In situ Hybridization Chain Reaction (HCR)

2.6.1 Sectioning

Tissue sections were collected as describe in Section 2.3.2 in series to spread the sections across many slides so multiple analyses could be performed on similar sections (Table 2.1). 10 µm thick sections were collected on Histobond+ slides (Mariesfeld) and stored at -20°C until further use.

For *in situ* hybridization, all operations were carried out under ribonuclease-free conditions by cleaning the lab bench and tools with RNase Zap (Invitrogen). The slides with tissue were thawed in 4% paraformaldehyde (Alfa Aesar) for 15 min to fix the tissue. The slides were transferred in 70% ethanol for 2 hours and permeabilized for 10 min in 8% SDS (Sigma-Aldrich). Lastly the slides were washed in 70% ethanol twice for 1 min each.

2.6.2 Detection stage

In this stage the probe sets carrying the DNA HCR initiators hybridize to the mRNA targets. A humidified chamber was prepared and pre-warmed to 37°C. The slides were dried by blotting on paper towels. The slides were pre-hybridized by an addition of 200 µl probe hybridization buffer (Molecular Instruments) and covered the slide with a parafilm coverslip for 10 min inside the humidified chamber. A probe solution was prepared by adding an optimized concentration of each of the probe mixture to 100 µl of probe hybridization buffer at 37°C (Table 2.18). The probes were ordered from Molecular Instruments (Table 2.18). The pre-hybridization solution was removed, and excess solution drained onto a paper towel. 50 µl of the probe solution was added onto the slides, except for the negative control slides which had 50 µl probe hybridization buffer only. The slides were covered with a glass coverslip. This was incubated overnight for 12-16 hours at 37°C inside the humidified chamber.

Probe wash buffer (Molecular Instruments) and SSCT (5X SSC [Sigma-Aldrich], 0.1% Tween-20 [Sigma-Aldrich]) were pre-warmed at 37°C. The slides were immersed in the probe wash buffer at 37°C to float off the coverslips. Excess probes were removed by incubating slides at

37°C for 15 min each in 75%, 25%, 50% and 25% of probe wash solution diluted in 5X SSCT. This was followed by a 15 min wash in 5x SSCT at 37°C and a 5 min wash in 5X SSCT at room temperature.

2.6.3 Amplification stage

In this stage the DNA hairpins penetrate the sample. The specifically bound probe pairs trigger self-assembly of a tethered fluorescent amplification polymer. The slides were dried by blotting the edges on a paper towel. The slides were pre-amplified in a humidified chamber for 30 min at room temperature by adding 200 µl amplification buffer and covered with a parafilm coverslip. The hairpin solutions (h1 and h2; [Molecular Instruments]) (Table 2.18) were prepared separately by heating 2 µl of each 3 µM stock (6pmol) at 95°C for 90 seconds and snap-cooling to room temperature in a dark drawer for 30min. The final hairpin mixture was prepared by adding the snap-cooled h1 hairpins and snap-cooled h2 hairpins to 100 µl of amplification buffer at room temperature. The pre-amplification buffer was removed. 50 µl of the hairpin mixture was added onto the experiment as well as the negative control slides. The slides were covered with a glass coverslip. The slides were incubated at 37 °C overnight (12hr-16hr) in a dark humidified chamber at room temperature.

To float off the coverslip the slides were immersed in 5X SSCT at room temperature. The excess hairpins were removed by incubating the slides in 5X SSCT at room temperature twice for 30 min and once for 5 min. To stain the nuclei, a 1 in 5000 dilution of Hoescht (Thermofisher) was added and incubated at RT in the dark for 10 min. For selected experiments, TrueBlack (Biotium) was used to quench the lipofuscin present in the human brain samples. The vial of the stock solution of TrueBlack, 20X in DMF, was preheated to 70°C for 5 minutes followed by diluting 1X in 70% ethanol and briefly centrifuged. The optimization of Trueblack was done by Jessica More (Honours student in the Hockman lab). Each slide was treated with 200 µl TrueBlack for 30 sec. After TrueBlack treatment the slides were rinsed in 1X PBS for 5 min. Slides were mounted by adding SlowFade (Invitrogen) and a glass coverslip.

2.6.4 Imaging

The slides were imaged using the LSM 880 Airyscan confocal microscope (Carl Zeiss, ZEN SP 2 software) using the 60X/40X objective. For imaging, the slides were subjected to lambda scanning which allows for detection of several fluorophores across a desired range of wavelengths. Lambda scanning also facilitated the detection of the background lipofuscin in the samples, which is present across most wavelengths. Lambda scanning/stack coupled with linear unmixing was used to separate the fluorescent signals from one another to clearly show the contribution of each fluorophore to the signal. Tile scanning, using the 40X objective, was used to capture all the cortex layers of the tissue.

Table 2.18: HCR probes and hairpins. The table shows the probes, probe concentration and the hairpins used for all samples. *Fabp7* and *Reln* analysis was performed by honours student, Jessica More, under my supervision.

Probes	Accession Number	Probe concentration	Hairpins used
<i>SLC17A7</i>	NM_020309.4	0.8 pmol	594/488
<i>GAD1</i>	NM_001650.7	1.6 pmol	647
<i>AQP4</i>	NM_001650.5	0.8 pmol	488
<i>RORB</i>	NM_006914.4	1.6 pmol	647
<i>TSHZ2</i>	NM_173485.6	1.6 pmol	488
<i>CLSTN2</i>	NM_022131.3	0.8 pmol	546
<i>CXCL14</i>	NM_004887.5	0.8 pmol	546
<i>FABP7</i>	CR457057.1	1.6 pmol	647
<i>RELN</i>	NM_005045.4	1.6 pmol	594

CHAPTER 3: RESULTS

The human temporal cortex tissue samples used were obtained during surgical treatment of epilepsy which involves the removal of a small area of the brain where the seizures originate. The tissue obtained in this study was the tissue located outside the epileptic focus, which is likely to be non-epileptic.

3.1 Visium Tissue Optimization

The Visium Tissue Optimization Protocol was conducted using RNA extracted from a 5-year-old flash frozen brain tissue fragment as the positive control RNA (See Table 2.1). The RNA has a RIN of 8.90 which indicates that the RNA was intact (Fig. 3.1 A). The RIN is a measure used to evaluate the quality of rRNA, serving as an indicator of the intactness of mRNA in a tissue. To optimize the permeabilization time the Visium Tissue Optimization Protocol was conducted using tissue sections obtained from a 15-year-old sample (See Table 2.1). The optimal permeabilization time was determined as 12 min, which is seen by the maximum fluorescence signal and high fluorescence peak over the 12 min section (Fig. 3.1 B). After the optimal permeabilization time was determined, the tissue samples (See Table 2.1) were sectioned onto Visium slides by first scoring the tissue to ensure the sections would fit in the capture areas on the Visium slides (see Fig. 3.1 D).

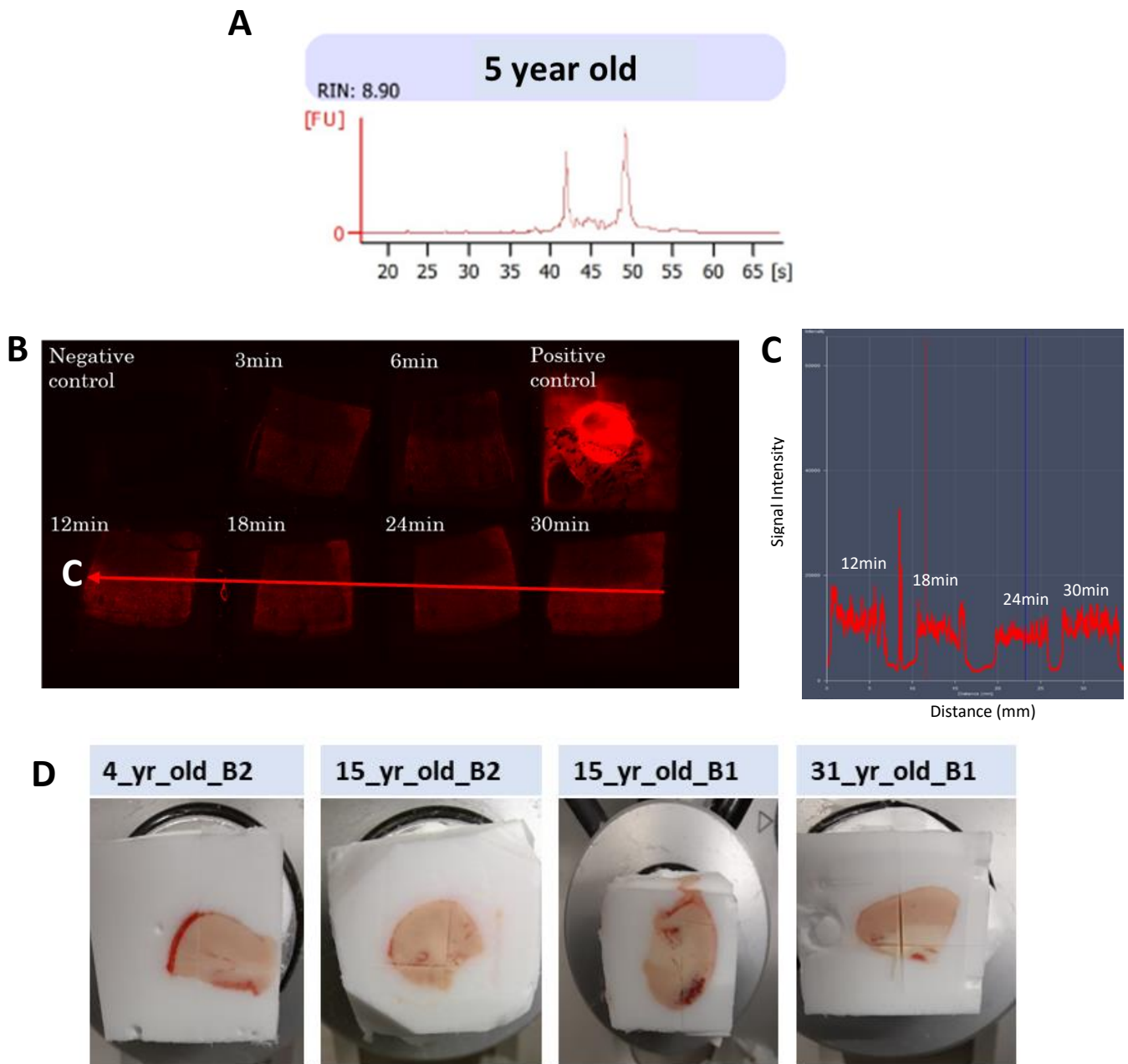


Figure 3. 1: Visium Tissue Optimization and sectioning

A) RNA quality assessment for the 5-year-old which was used as the positive control in B. B) The permeabilization optimization time course conducted on the 15-year-old tissue sections showing the fluorescent signal across all sections. C) Analysis of the fluorescence output across the bottom row in B (red arrow) shows the highest fluorescence output over the 12 min section D) Images of the OCT blocks of the samples used in this study, showing the scoring done to each section.

3.2 Spatial transcriptomics using 10X Genomics Visium Spatial Gene expression.

Spatial transcriptomics allows us to measure all the gene activity in a tissue section and to map the location of this activity. This helps us understand how a cells' localization in the tissue influences its gene expression profile. In this study we used the 10X Genomics Visium Spatial Gene Expression protocol which provides the gene expression information while retaining the organization of the tissue. Two technical replicates Visium libraries were prepared for each sample (see Table 2.1)

After frozen tissue sections were placed onto the Visium slides, the slides were H&E stained and imaged (see Fig. 3.6). This was followed by conducting the Visium protocol to generate sequencing libraries. The Visium protocol involved quality control (QC) checks. The initial quality control check was done at the post cDNA amplification stage where the cDNA libraries were run on the Agilent Tapestation to assess the fragment size distribution (Fig. 3.2 A; Supp. Fig. 1) and their concentration determined (Table 3.1). At this stage, the samples all showed discreet peaks greater than 500 bp, except for 4_year_old_T2 which had two joined peaks (Supp. Fig. 1). The cDNA Quantification of the 1st Visium experiment showed a more consistent concentration between replicates compared to the 2nd Visium experiment (Table 3.1). The total number of Sample Index PCR cycles was optimized based on the carry forward proportion of the total cDNA yield (ng) (Table 3.1). Despite the differences in the cDNA yield in the two Visium experiments, the same Sample Index PCR cycle number of 14 was used (Table 1). This was done so that the two experiments are comparable. The last QC was done after PCR indexing/library construction. Again, the post library construction QC of the 1st Visium experiment had a more consistent concentration between replicates compared to the 2nd Visium experiment (Table 1). The final libraries were also run on the Agilent Tapestation to assess the fragment size distribution (Fig. 3.2 B; Supp. Fig. 2). For the final library, 10X Genomics recommends a library with a single discrete peak with a bell curve at 300-600bp in size and an average library size of approximately 450bp (Suppl. Fig. 2). All the samples had an average library size between 350 and 550 bp post PCR indexing (Fig. 3.2 B; Supp. Fig. 2).

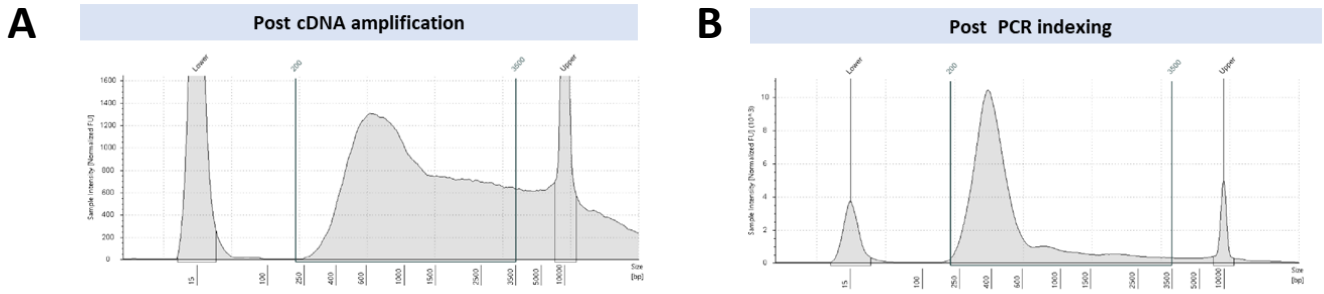


Figure 3.2: Quality control during the Visium process

A) Tapestation trace indicates the cDNA fragments size distribution after cDNA amplification for the 31-year-old (T1_B1). B) Tapestation trace which shows the fragment size distribution after PCR Indexing. Lower and Upper markers used are included in each trace for the 31-year-old (T1_B1). The region that was set for the calculation of cDNA concentration is indicated by the vertical green lines on either side of the library trace.

Table 3.1: Quality Control metrics for the 1st (B1) and 2nd (B2) Visium experiments. The Visium libraries were quantified after cDNA amplification and post library construction.

Age	cDNA QC & Quantification ng/ul	cDNA yield(ng)	10 µl carry forward for step 4.1(ng)	Post library construction QC (ng/µl)
15-year-old_T1_B1	18.7	654.5	187	41
15-year-old_T2_B1	17	595	170	42.3
31-year-old_T1_B1	17.8	623	178	49.2
31-year-old_T2_B1	24.4	854	244	49.9
4-year-old_T2_B2	51.0	1928	482	18.9
4-year-old_T1_B2	33.5	1064	266	21.8
15-year-old_T2_B2	7.16	396.8	99.2	6.67
15-year-old_T1_B2	30.3	2096	524	45.7

3.3 Library sequencing and alignment

Before sending the libraries for sequencing, the H&E images were processed using the 10X Genomics Loupe Browser 4.0 Visium Manual Alignment Wizard to determine the required sequencing depth by calculating the number of spots that were covered by the tissue sections (Table 3.2). Each Visium slide has 4 capture areas, and each capture area contains 5000 barcoded spots. All the tissues samples covered more than 3000 barcoded spots (Table 3.2).

Table 3.2: The number of barcoded spots on the Visium capture area that were covered by the tissue for each sample in the 1st (B1) and 2nd (B2) Visium experiments.

Samples	Number of Spots Covered
15-year-old_T1_B1	3636
15-year-old_T2_B1	3652
31-year-old_T1_B1	3504
31-year-old_T2_B1	3409
4-year-old_T2_B2	3711
4-year-old_T1_B2	3168
15-year-old_T2_B2	3314
15-year-old_T1_B2	3186

The libraries were sent for sequencing and the raw data was analysed using the 10x Genomics *Spaceranger count* pipeline. *Spaceranger count* was used to perform alignment (using human reference genome), tissue detection, fiducial detection and barcode/UMI counting. While the sequencing saturation and % mapped to the transcriptome of the 2nd Visium experiment was higher than the 1st Visium experiment, all the other sequencing metrics were lower (Table 3.3). The mean reads per spot is dependent on the sample type and RNA. 10X Genomics recommends a value of 50 000 mean reads per spot and all the samples had a mean reads per spot > 50 000 (Table 3.3). According to 10X Genomics the median reads per spot can vary by sample, this is noticed in our samples. The 1st Visium experiment had a higher median reads per spot values when compared to the 2nd Visium experiment. The 1st Visium experiment also showed more consistent median per spot values than the 2nd Visium experiment. The Median UMI counts per spot is dependent on the RNA quality, tissue type and sequencing depth. High values were seen in the 1st Visium experiment and very low values in the 2nd Visium experiment, with 4_year_old_T1_B2 showing the lowest value (Table 3.3). These low values could be an indication of low library complexity or quality. The total genes detected is dependent on the sample type, RNA quality and sequencing depth and the 2nd Visium experiment had much lower values compared to the 1st Visium experiment.

Table 3.3: Web summary results from for the 1st (B1) and 2nd (B2) Visium experiments showing the metrics that can be used to assess the success of the experiments.

Sample	Number of spots under tissue	Mean reads per spot	Median genes per spot	Total genes detected	Median UMI Counts per Spot	% mapped to Transcriptome	Sequencing Saturation (%)
15-year-old_T1_B1	3 651	90 727	2 858	23 144	6 206	48.7	80.5
15-year-old_T2_B1	3 723	80 075	2 753	22 937	6 087	53.0	80.5
31-year-old_T1_B1	3 574	112 963	2 636	22 367	5 356	43.2	82.2
31-year-old_T2_B1	3 376	83 628	2 736	22 821	5 668	45.4	78.6
4-year-old_T2_B2	3 711	125 949	673	19 862	905	59.1	98.0
4-year-old_T1_B2	3 168	80 631	414	19 084	539	69.9	98.4
15-year-old_T2_B2	3 314	124 151	984	20 093	1 374	59.2	95.9
15-year-old_T1_B2	3 186	75 179	580	19 249	755	67.7	96.8

3.4 Estimating the number of nuclei per spot

The number of nuclei per spot is needed for downstream analysis. The 10x Genomics Visium Spatial Gene Expression slide has 5000 spots that are each 55 µm in diameter. This means that each spot can capture 1 to 10 cells on average. Vistoseg (Tippani et al., 2021) was used for segmentation, analysis, and visualization of Visium HE images to obtain an estimation of the nuclei number per spot (Fig. 3.3 A-B). The number of nuclei obtained for the 1st Visium experiment is higher than the 2nd Visium experiment (Table 3.4). The average nuclei per spot for the 1st Visium experiment was higher than the 2nd Visium experiment (Table 3.4). The Vistoseg segmentation results showed that some of the spots contained more than 10 nuclei. Once Vistoseg was performed on all the samples, the average number of nuclei per Visium location per sample was obtained. Followed by adding all the averages and dividing by 8 (all the samples) and this was used as a hyperparameter (N_cells_per_location) during spatial mapping of cell types using cell2location.

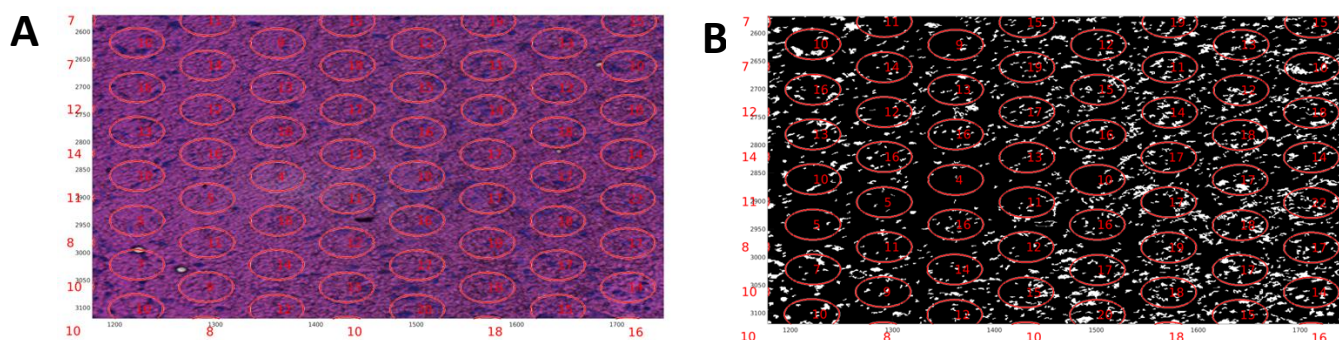


Figure 3.3: Segmentation to estimate the number of nuclei per spot in the Visium samples.

A) H&E image segmentation and number of nuclei found in each spot for 31-year-old_T2_B1 sample. B) The number of nuclei per spot after segmentation. The red circles indicate the position and size of the spots (55 μ m) on the Visium Spatial Gene Expression slide. The red numbers inside the circles indicate the number of nuclei found in that spot.

Table 3.4: Shows the total number of nuclei for all the spots covered by each tissue samples in the 1st (B1) and 2nd (B2) Visium experiments.

Sample (Age)	Total Number of nuclei for all the spots	Average nuclei per spot
15-year-old_T1_B1	50434	10.14
15-year-old_T2_B1	55148	11.06
31-year-old_T1_B1	43704	8.79
31-year-old_T2_B1	56828	11.42
4-year-old_T2_B2	25015	5.02
4-year-old_T1_B2	13125	2.63
15-year-old_T2_B2	15038	2.81
15-year-old_T1_B2	13989	3.02

3.5 Spatial mapping and abundance of 54 cell types across the Visium samples

After Spaceranger *count* and Vistoseg analysis, the Cell2location tool was used to spatially map the brain cell types by integrating the Visium count matrices with existing single nucleus (sn)RNAseq datasets which had been analysed by Christina Steyn (Masters' student in the Hockman lab) (Fig. 3.4A). This reference dataset had been annotated using the current published human middle temporal gyrus cell atlas as a guide (Hodge et al., 2019), in which each cell type is annotated with 1) a major cell class (e.g. Exc for excitatory neuron), 2) the cortical layer the cell is associated with (e.g. L2 for layer 2), 3) a subclass marker gene and 4) a cluster-specific marker gene.

Spatial mapping was performed using cell2location to estimate which combination of cell types could have given rise to the mRNA signal in each spot in the Visium data. Cell2location

was able to estimate the relative abundance of 54 cell types across the different samples (Fig. 3.4). Unsurprisingly, non-neuronal cell types: astrocytes, oligodendrocytes and endothelial cell were amongst the most abundant cell types, oligodendrocytes being the most abundant. For the neuronal cell types, the excitatory neuron subtype *Exc L2 LAMP5 LTK* and the inhibitory neuron subtype *Inh L1-2 VIP TSPAN12* were the most abundant (Fig. 3.4). *Exc L3-5 RORB COL22A1* and *Exc L3-5 RORB FILIP1L* were amongst the least abundant subtypes in the excitatory neuron category (Fig.3.4). Most of the inhibitory neuron subtypes were less abundant than the other cell subtypes.

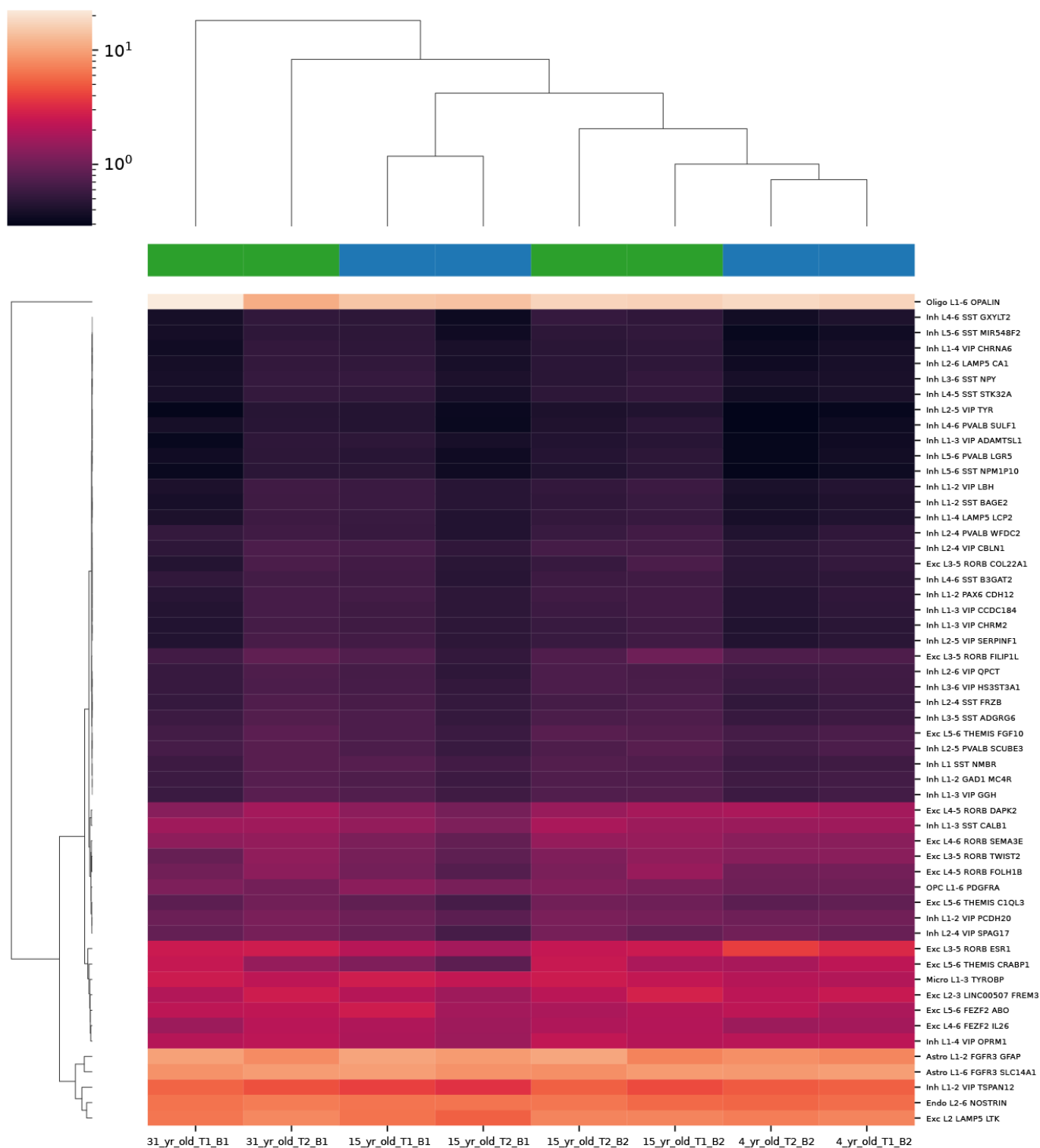


Figure 3.4: Spatial mapping of cell types in the human temporal cortex.

Estimated cell abundance of 52 cell types across all Visium samples. Shown is a heatmap with the colour indicating the relative cell abundance of cell types (rows) across the different age groups (columns). The brighter colour indicates higher cell abundance.

Cell2location was then used to map the cell types to their locations in the tissue sections. The reference cell types mapped to the expected tissue regions (Fig 3.5; Supp. Fig. 3). The astrocytes were highly enriched in layer 1 of the cortex and the white matter (WM) (Fig. 3.5). Microglia and Oligodendrocytes precursor cells (OPCs) were found mostly in layer 1 and WM but were also present in a few spots in the cortical layers. Oligodendrocytes were found mostly in the WM. Most of the excitatory neurons were layer enriched and mostly found from layers 2-6 (Fig. 3.5). The inhibitory neurons were not as layer-specific and were found distributed in all the cortex layers, with very low levels in the WM.

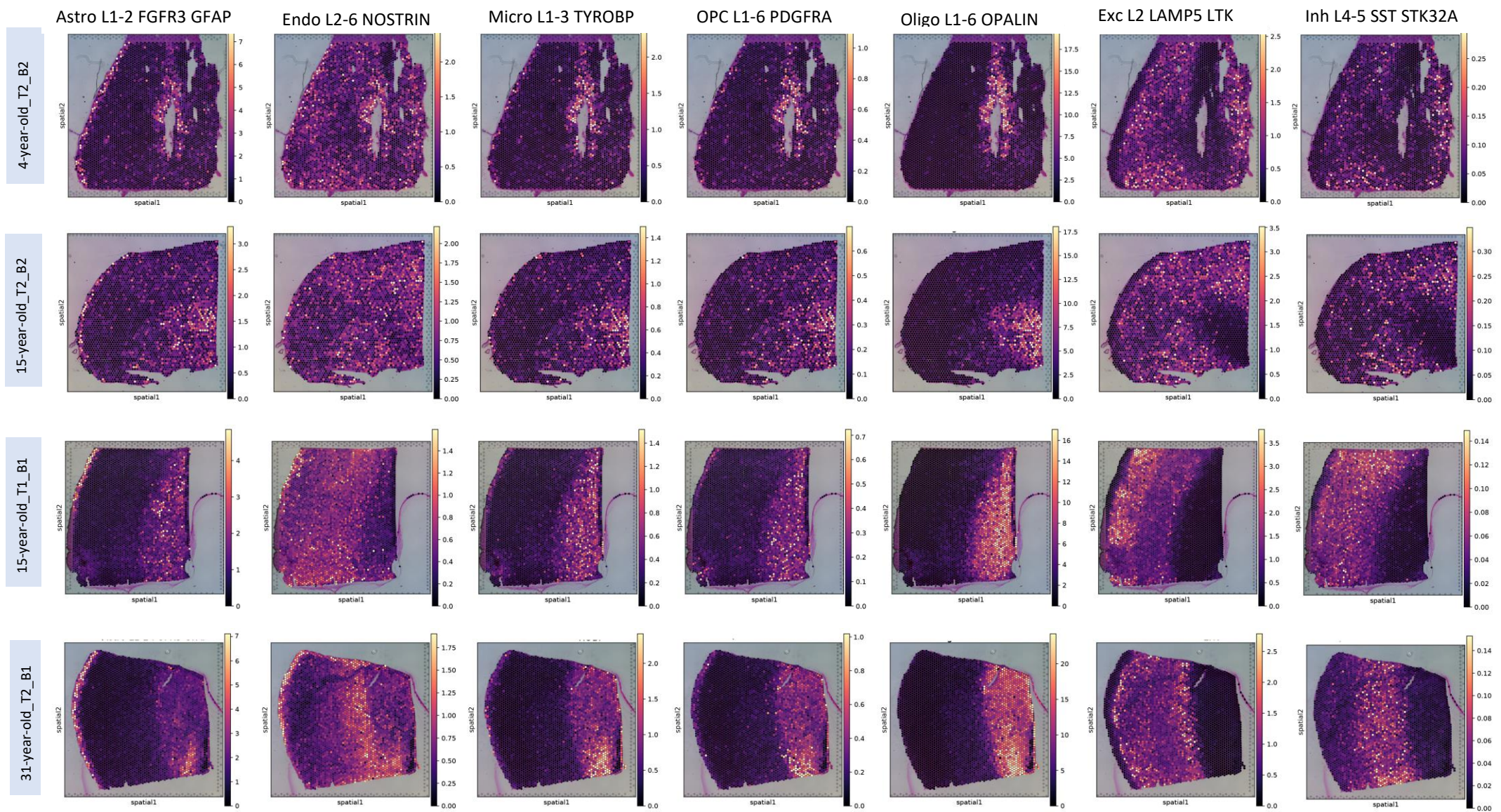


Figure 3. 5: Spatial mapping of cell types in the human temporal cortex.

Estimated cell type abundance plots for a selection of the major brain cell types in a representative section (technical replicate) for each patient sample. Brighter signal indicates a higher abundance of the cell type

3.6 Identification and verification of layer-specific cell types

Using cell2location to spatially map multiple cell types together, the results captured the layer-specific spatial patterns within the temporal cortex tissue (Fig. 3.6; Supp. Fig. 4) (see github script `Fine_celltypes.py`). This supported the annotation of the cortical layers and the white matter across the Visium H&E reference images (Fig. 3.6 A-B). In all samples, the astrocyte subtype *Astro-L1-2 FGFR3 GFAP* was highly expressed in layer 1 as well as the WM (Fig. 3.6 C). Similarly, as expected, the oligodendrocyte subtype *Oligo L1-6 OPALIN* was seen in the WM in all the samples. The 1st Visium experiment clearly showed various excitatory subtypes marking cortical layers 2-6 in both the paediatric and adult samples, however this was not seen as clearly in the 2nd Visium experiment (Fig 3.6C). It is important to note that the overall quality of the 2nd Visium experiment was not ideal based on the genome alignment results (see Table 3.3). Strikingly, *Exc L3-5 RORB ESR1* clearly marked cortical layer 4 in the samples from the 1st Visium experiment (Fig 3.6 C; Supp. Fig. 4 C). The distribution of *Exc L2 LAMP5 LTK* in the paediatric samples is widely dispersed and becomes more enriched in the upper layer 2 in the adult sample (Fig. 3.5; Fig 3.6 C; Supp. Fig. 4 C). This cell type is also present in layer 6 with increasing abundance in the adult samples (Fig. 3.6). Most Inhibitory neuron subtypes were sparsely distributed throughout the cortical layers, with *Inh L1-4 LAMP5 LCP2* showing a greater association with cortical layer 1 than other inhibitory neuron sub-types in paediatric and adult samples (Fig 3.6 C; Supp. Fig. 4 D).

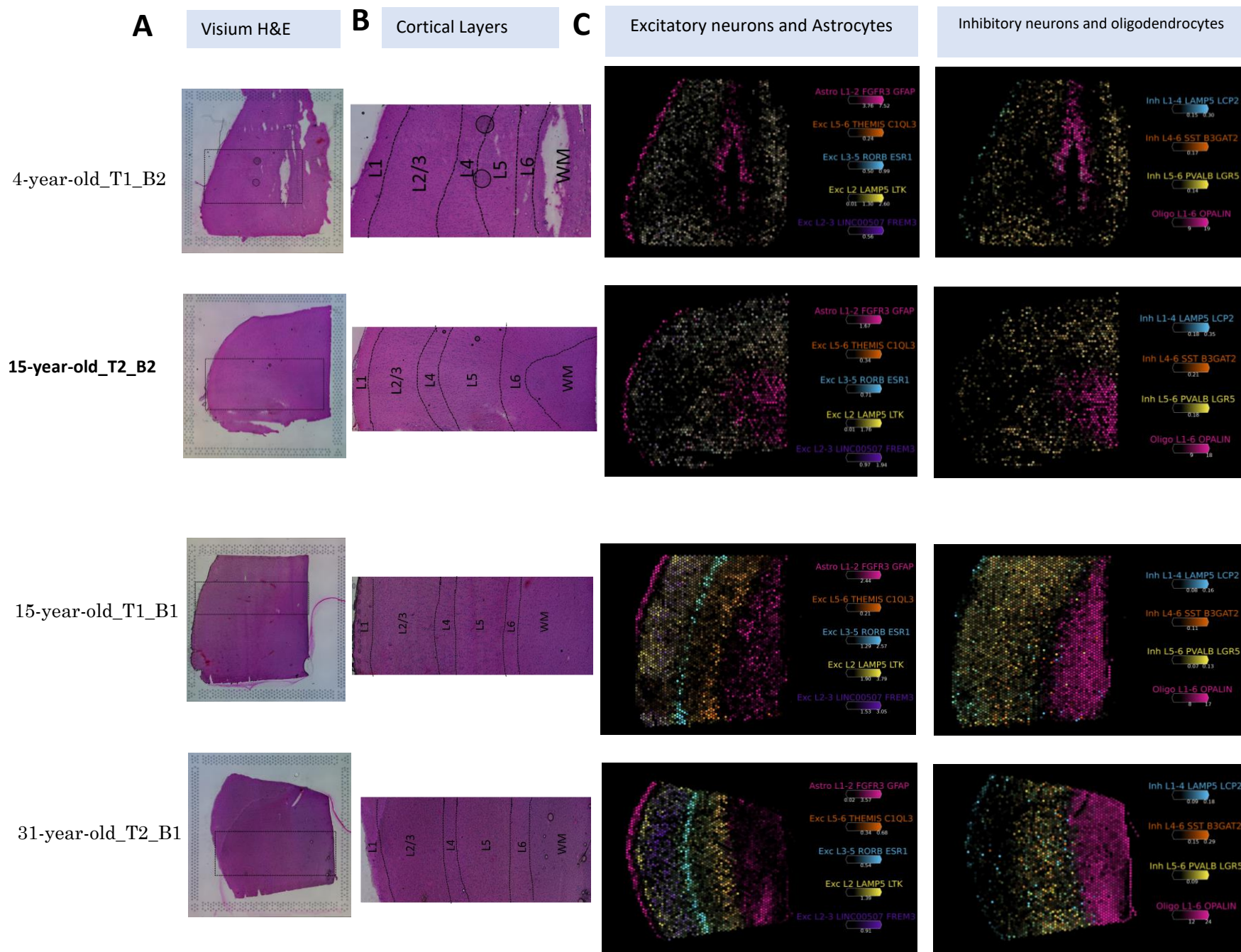


Figure 3.6: Spatial mapping of layer-specific cell types.

A) H&E reference images for selected technical replicate temporal cortex sections generated during Visium Library preparation. The box indicates the region of the section that is magnified in the adjacent panel. B) Magnified H&E images annotated with the estimated boundaries of cortical layers 1-6 and the white matter (WM) marked by black dotted lines. C) Estimated cell type abundances (colour intensity) for a selection of cell types that clearly denote the cortex layers, especially in the 15-year-old-T1-B1 and 31-year-old-T2-B1 samples.

3.7 Identifying cellular compartments using non-negative matrix factorisation

To be able to understand tissue organisation and detect cellular interactions or compartments, non-negative matrix factorization (NMF) was performed using the cell abundance estimates from cell2location. NMF was trained for a range of factor settings (5 – 30) and factor setting of 15 was chosen as a balance to capture distinct cellular compartments and co-localisation of cell types (Kleshchevnikov et al., 2022). The NMF weights for each of the 15 identified cellular compartments were visualised across the Visium samples to expose their spatial distribution (Fig. 3.7; Supp. Figs 5-6) and the cell types that were associated with each compartment were determined (Fig. 3.7 B).

Factors 7, 8, 9, 12 and 13, were largely associated with non-neuronal cells, most of the which were enriched in layer 1 and the WM, except for factor 8, which was found in the cortical layers only (Fig. 3.7 A-B). Factor 7 was associated with endothelial cells, which were more prominent in layer 1 across all the sample but present across the rest of the cortical layers and the WM (Fig. 3.7 A-B; Supp. Figs. 5-6). Factor 8 distinctly contained *Astro L1-6 FGFR3 SLC14A1* which was enriched in layer 1 and found sparsely in the rest of the cortical layers excluding the WM across all samples (Fig 3.7 A-B; Supp. Figs. 5-6). Factor 9 distinctly contained *Astro L1-2 FGFR2 GFAP* which marked layer 1, WM and a few spots in the rest of the cortex. It showed a change in the distribution with age, it has a higher abundance in the WM of the paediatric samples compared to the adult samples (Fig 3.8A). Factor 12, which was confined to the WM, was strongly associated with *Oligo L1-6 OPALIN*. Factor 13 showed the colocalization of microglia and OPCs in layer 1 and the WM across all the samples, with microglia showing a higher abundance than OPCs (Fig. 3.7 A -B).

Factors 4, 6, 10, 11 and 14 were associated with excitatory neurons, recapitulating the layers of the cortex. Factor 6, which contains *Exc L2-3 LINC00507 FREM3*, is found in the broad region of layer 2/3 in all the samples (Fig. 3.7 A-B; Supp. Figs. 5-6). Factor 11, which distinctly contained *Exc L2-3 Lamp5 LTK*, was found in layer 2/3 and layer 6 across all the samples (Fig. 3.7 A-B; Fig. 3.8B; Supp. Figs 5-6). Factor 10 was found in the distinct region of the thin layer 4 in the 1st Visium experiment. In the 2nd Visium experiment this factor was more abundant towards the middle of the cortex where layer 4 is located but was not as clearly layer restricted (Fig. 3.8 A). This could be due to the lower quality of the samples or the medical condition in these patients. This factor was associated with *Exc L3-5 RORB FILIP1L*, *Exc L3-5*

RORB TWIST2, *Exc L3-5 RORB ESR1* and *Exc L4-5 RORB DAPK2*. Factor 2, which showed an overlap with layers 4-5 in the 1st Visium experiment and expanded into layer 6 in the 2nd Visium experiment, was associated with a colocalization of *Exc L4-5 RORB FOLH1B*, *Exc L4-5 RORB DAPK* and *Exc L4-6 RORB SEMA3E* (Fig. 3.7 A-B; Supp. Fig. 5-6). Factor 4 was associated with *Exc L5-6 THEMIS CRABP7* and marked layer 5/6 across all samples. Factor 14 distinctly contained *Exc L5-6 FEZF2 ABO* and marked layer 6 across all samples (Fig. 3.8 B).

Inhibitory neurons made up factors 0, 1, 3 and 5, and were not as layer defined as excitatory neurons (Fig. 3.7 B). *Inh L2-4 VIP SPAG17* was the dominant cell type in Factor 0 which was spread throughout the cortex and WM but prominent in layer 1 in most of the samples (Fig 3.7 A). Interestingly factor 1 was restricted to cortical layer 5-6 in the adult sample, but not in the rest of the samples (Fig. 3.8 A). The corresponding highest abundance cell type, *Inh L1-2 VIP TSPAN12* (Fig. 3.7 B), was more widespread than its expected spatial location (Fig. 3.8 A). In Factor 5, *Inh L3-6 SST NPY* was dominant and sparsely spread through the cortex, with a few spots in the WM (Fig. 3.7 A; Suppl. Fig 5-6). Factor 3 contains inhibitory and excitatory neurons at different abundance level, with inhibitory neurons being more dominant (Fig. 3.7 B). These cell types are sparsely spread throughout the cortex and less in L1, they are more prominent in L4/5/6 (Fig. 3.7 A; Suppl. Fig 5-6).

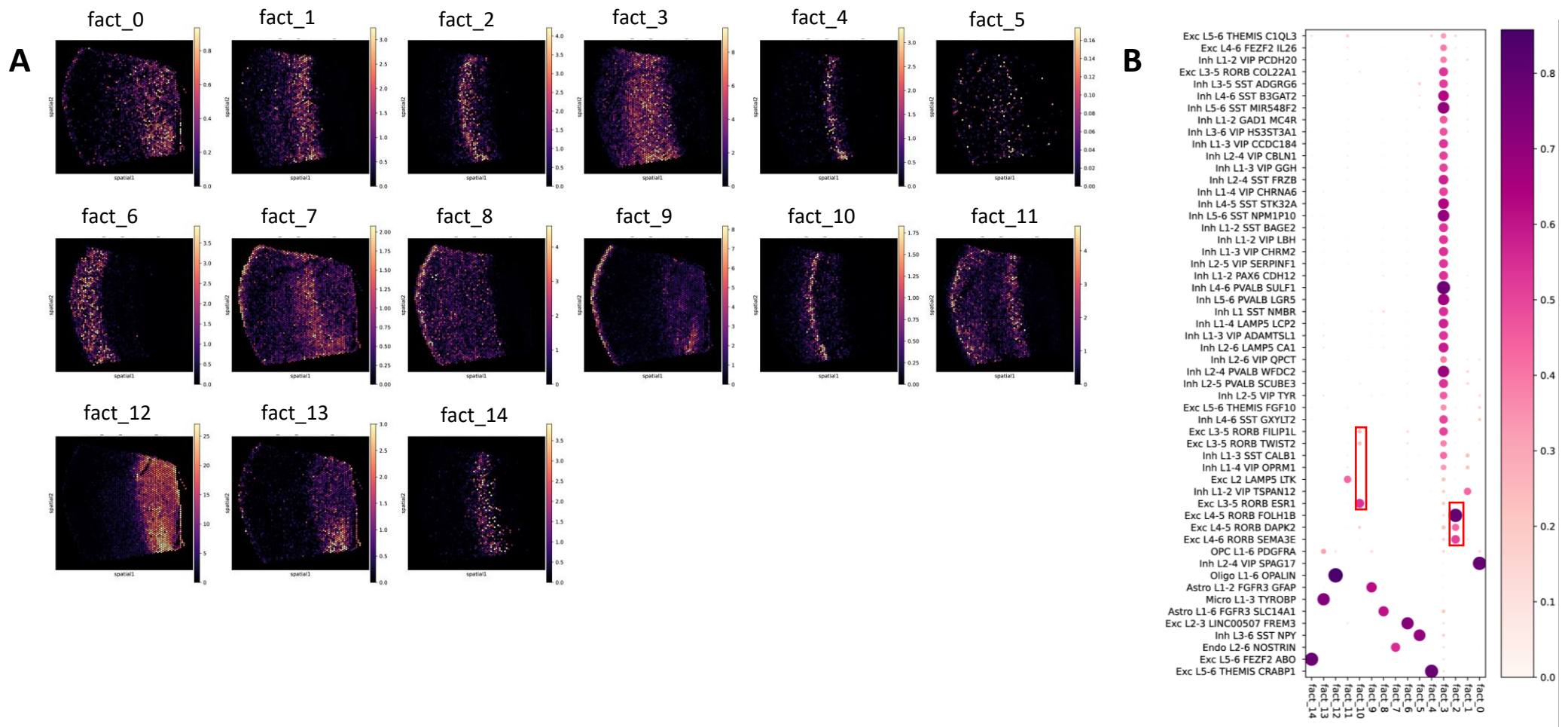


Figure 3. 7: Mapping the cellular compartments using non-negative matrix factorisation (NMF).

A) Spatial plots showing the NMF weights in a representative 31-year-old sample. B) Dot plot showing the NMF weights of the cell types (rows) across each of the NMF factors (columns), which corresponds to tissue compartments. Red boxes indicate cell types that co-locate.

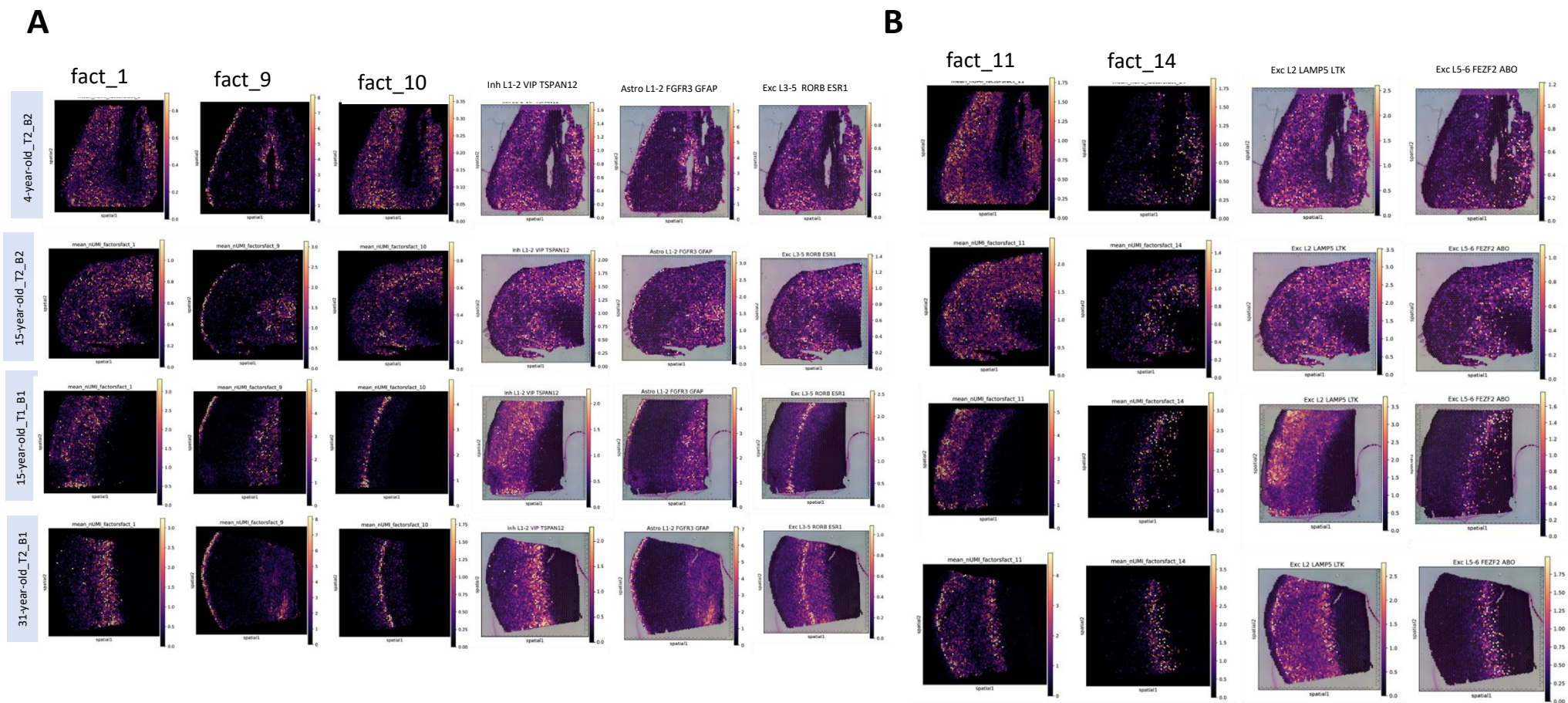


Figure 3. 8: Mapping the cellular compartments using non-negative matrix factorisation (NMF).

A) Spatial plots showing selected NMF factors that become region/layer specific with age alongside spatial plots of their defining cell types.

B) Spatial plots showing selected NMF factors that have a similar spatial distribution across samples alongside spatial plots of the corresponding cell types.

3.8 Leiden clustering

Leiden clustering was used to identify tissue regions that differ in their cell composition using the cell abundance estimates from cell2location. Leiden clustering differs from NMF because it shows the distribution of cell types across distinct, non-overlapping regions. The K-nearest neighbour (KNN) was computed using $n_neighbours = 21$. This was followed by Leiden clustering using a resolution of 0.8, resulting in a total of 20 region clusters (Fig. 3.9). The number of KNN neighbours and Leiden resolution used was based on the clustering that matched the predicted anatomical structure of the tissue samples. The clustering was done on the integrated Visium object containing all the Visium libraries so that the region identities can be comparable across samples.

A joint integrated UMAP plot of the data from the two Visium experiments was computed and showed good sample integration (Fig. 3.9 A-B). The UMAP was divided into two major groups, consisting of spots from both Visium experiments (Fig. 3.9 A), with 6 clusters on the right side and 15 clusters on the left side of the plot (Fig. 3.9 B). Individual UMAP plots for each sample showed the clusters present in each sample and the spatial plots showed the location of the clusters across each tissue section (Supp. Fig. 7). Spatial plots of the clusters revealed that the group on the right consisted of WM clusters, whilst the group on the left consisted of grey matter clusters (Fig. 3.9 C-F; Supp. Fig. 7). The WM clusters (6, 9, 10, 12, 15, 16) were made up of spots from all the samples (Table 3.5). Some of the clusters were sample/batch specific. For example, cluster 19 was predominantly present in one adult sample (31_year_old_T2_B1) compared to other samples, where cluster 19 consisted of very few or no spots (Table 3.5). Cluster 20 was made up of spots from the 1st Visium experiment only (Table 3.5). Cluster 17 is present in all the samples except for the two replicates of the 4-year-old (Table 3.5).

To investigate if the clusters corresponded with distinct cortical layers, the spatial layout of individual clusters was plotted separately for each Visium sample (Fig 3.9 C- F; Supp. Fig. 8). Additionally, the cell types associated with each cluster was determined (Fig 3.9 G). To further understand the genes that differentiate the clusters from each other, the Seurat function *FindAllmarkers* was used to identify the top 20 differentially expressed genes per cluster (Fig 3.10).

Clusters 6, 9, 10, 12, 15, 16, which mapped onto the white matter in all samples (Fig 3.9 F, Supp. Fig. 8 D), were dominated by varying levels of non-neuronal cell types including *Micro L1-3 TYROBP*, *OPC L1-6 PDGFRA*, *Oligo L1-6 OPALIN*, *Astro L1-2 FGFR3 GFAP* and *Endo L2-6 NOSTRIN* (Fig 3.9 G). These clusters also included inhibitory neurons at low levels, with *Inh L1-2 VIP SPAG17* being the most highly abundant. Interestingly, these six white matter clusters can be defined by the abundance levels of these various cell types. Cluster 16 showed the highest abundance levels of both non-neuronal and inhibitory neuron cell sub-types compared to the other clusters, while cluster 15 showed the lowest. Cluster 10, 12 and 15 had low levels of *Micro L1-3 TYROBP*, *OPC L1-6 PDGFRA*, *Oligo L1-6 OPALIN*, *Astro L1-2 FGFR3 GFAP* and *Endo L2-6 NOSTRIN* (Fig 3.8 G), while the levels of *Astro L1-2 FGFR3 GFAP* and *Endo L2-6 NOSTRIN* were higher. This is further validated by the expression of astrocyte and endothelial gene markers such as *TMSB10*, *CST3* and *GPM6A* at high levels in these clusters. These clusters also have higher levels of neuronal genes, such as *NRGN* and *TUBB2A*, compared to the other clusters. Cluster 6, 9 and 16 have higher levels of *Micro L1-3 TYROBP*, *OPC L1-6 PDGFRA* and *Oligo L1-6 OPALIN*. This corresponds with the expression of oligodendrocyte gene markers such as *PLP1*, *MBP*, *CNP*, *SOX2-OT* and *RNASE1* in these clusters (Fig 3.10 C). This cluster also expresses *PTGDS* gene which is expressed in both oligodendrocytes and astrocytes according to the Allen Brain Map dataset (Bakken et al., 2021).

Cluster 7 and 14 marked layers 1/6 and layers 1-5 respectively across all samples (Fig. 3.9 C; Supp. Fig. 8 A). The 4-year-old samples had more spots that belonged to cluster 7 compared to other samples. Both clusters are associated with *Astro L1-2 FGFR3 GFAP*, *Astro L1-6 FGFR3 SLC14A1* and *Endo L2-6 NOSTRIN*, along with lower levels neurons (Fig 3.9 G). Cluster 7 was highly enriched in layer 1, with a few spots found in layer 6 (Fig 3.9 C). In 4-year-old_T2_B2 cluster 7 is highly enriched in L1/6 but also seen across the cortical layers (Fig 3.9 C). Cluster 7 was associated with known astrocyte markers such as *GFAP*, *AQP4* and *S100B* (Fig 3.10 A). Endothelial cell associated genes, such as *VIM* and *SPARC*, were also identified in cluster 7 (Fig 3.10 A). In correspondence with Maynard et al., 2021, who identified *FABP7* as a layer 1 marker, *FABP7* was identified in cluster 7, where it was enriched in the 31-year-old compared to the rest of the samples (Fig. 3.10 A). Cluster 14, which had a more dispersed distribution, was dominated by *Astro L1-6 FGFR3 SLC14A1* but also contains many neuronal subtypes (Fig.

3.9 G). Cluster 14-associated genes included *ENC1*, which was expressed in the 15- and 31-year-old only (Fig. 3.10 A). *ENC1* was previously identified to be enriched in layer 2 (Maynard et al., 2021). Other genes marking this cluster were *OLFM1* and *NRGN*, which are expressed by neurons in the Hockman snRNA-seq data as well as publicly available datasets (Karlsson et al., 2021).

Cluster 17 was found in cortical layers L2/3 in all the samples except for the 4-year-old, in which it is absent (Table 3.5; Fig. 3.9 D; Supp. Fig. 8 B). The 15-year-old samples from the 1st Visium experiment had the most spots belonging to this cluster compared to the rest of the samples (Table 3.5). This cluster had high abundance levels of *Exc L2 LAMP5 LTK* and *Exc L2-3 LINC00507 FREM3* cell sub-types (Fig. 3.9 G) and was clearly associated with known layer 2-3 markers such as *ENC1* (Maynard et al., 2021) and *LAMP5* (Hodge et al., 2019) (Fig. 3.10 B). This cluster also expressed *HOPX* and *PCDH8*, which are expressed by neurons in publicly available datasets (Bakken et al., 2021; Karlsson et al., 2021) and are enriched in layer 2 (Maynard et al., 2021).

Cluster 20 was associated with cortical layer 4 and was seen only in the samples from the 1st Visium experiment (both replicates of the 15-year-old and only 1 replicate of the 31-year-old) (Table 3.5; Fig 3.9 D; Supp. Fig. 8 B). This cluster had high abundance levels of *Exc L3-5 RORB ESR1*, *Exc L3-5 RORB TWIST2*, *Exc L3-5 RORB FILIP1L*, *Exc L4-5 RORB DAPK2* and *Exc L4-6 RORB SEMA3E*, (Fig. 3.9 G). It was also associated with inhibitory neurons at high levels such as *Inh L1-2 VIP TSPAN12* and *Inh L1-4 VIP OPRM1*. The genes associated with this cluster included known markers for layer 4 (*RORB*) (Maynard et al., 2021) and layer 5 (*TSHZ2* and *PCP4*) (Tasic et al., 2018; Maynard et al., 2021) (Fig 3.10 B). *NREP* was another marker for this cluster that was highly expressed in the 15-year-old from the 1st Visium experiment. *NREP* is expressed in the excitatory and inhibitory neurons in the Allen Brain Map dataset, and it is seen in layer 4/5 in Maynard et al. 2021's SpatialLIBD dataset. *TSHZ2* has been previously identified as a marker for layer 5-6 in the adult mouse brain (Tasic et al., 2018) and in this study, it was more highly expressed in the 31-year-old sample compared to the 15-year-old samples. *TSHZ2* is expressed in the excitatory and inhibitory neurons in the Allen Brain Map dataset.

Cluster 18 marked cortical layer 5/6 and this was seen in all the samples (Fig 3.9 E; Supp. Fig. 8 C). Even though the samples from the 2nd Visium experiment had very few spots in this cluster, the spots were in layer 5/6 (Table 3.5). This cluster had high abundance levels of *Exc*

L5-6 FEZF2 ABO and *Exc L5-6 THEMIS CRABP1*. Genes associated with this cluster included markers for layer 5/6, such as *CLSTN2* (Zhou et al., 2020), *KRT17* (Maynard et al., 2021) and *SEMA3E* (Watakabe, 2009)(Fig 3.10 B). Interestingly, *CLSTN2* and *SEMA3E* showed higher expression in the 4-year-sample compared to the rest of the samples. *CLSTN2* is expressed by excitatory and inhibitory neurons and *SEMA3E* is highly expressed by excitatory neurons in the Allen brain dataset. Interestingly, *KRT17* is more strongly expressed in the 15-year-old from the 1st Visium experiment compared to the rest of the samples. (Fig. 3.10 B). *KRT17* is expressed by excitatory neurons (Bakken et al., 2021).

Clusters 0,1,2,3,4,5,8,11,13 and 19 were associated with varying levels of all cell types. With cluster 3, 8 and 13 showing the lowest levels, the 2nd Visium experiment had more spots belonging to these clusters compared to the 1st Visium experiment. Cluster 11 had the highest levels of all the cell types. These clusters were distributed across the cortex in most samples, except for cluster 0 and 1 which were highly abundant in layers 4-6 in all the samples except for 31-year-old-T1-B1 where it was distributed from layers 2-6 (See Supp. Fig. 7). Cluster 1 was distributed across the cortex in all the samples except for 31-year-old_T1_B1 where it was in layer 2/3. In the 1st Visium experiment cluster 4 is mostly seen in layer 1/2/3/4 and in the 2nd Visium experiment it is distributed across the cortex and the WM. Cluster 5 is distributed across different layers in all the samples and had a bias towards layer 5/6 in 15_year_old-T1-B2 and 15-year-old_T1_B1. Cluster 19 shows a high abundance of *Exc L2-3 LINC00507 FREM3* compared to the rest of the cell types found in this cluster. It is also distributed in layer 2/3 across all the samples and absent in 4-year-old_T2_B2.

Overall, both NMF and Leiden clustering analyses were able to provide information on the spatial location of cell types within the temporal cortex. The NMF analysis provided high resolution clusters made up of cell types that are distinctly found at a particular location/layer within the tissue, while also exposing some co-localising cell types within the layers of the cortex, such as colocalising excitatory neuron sub-types. Leiden clustering revealed more broad tissue clusters that varied in the abundance level of each cell type, and exposed which cell type is most highly abundant in each region. The Leiden clustering was also used to determine the gene expression patterns profiles associated with the cell types in each tissue region.

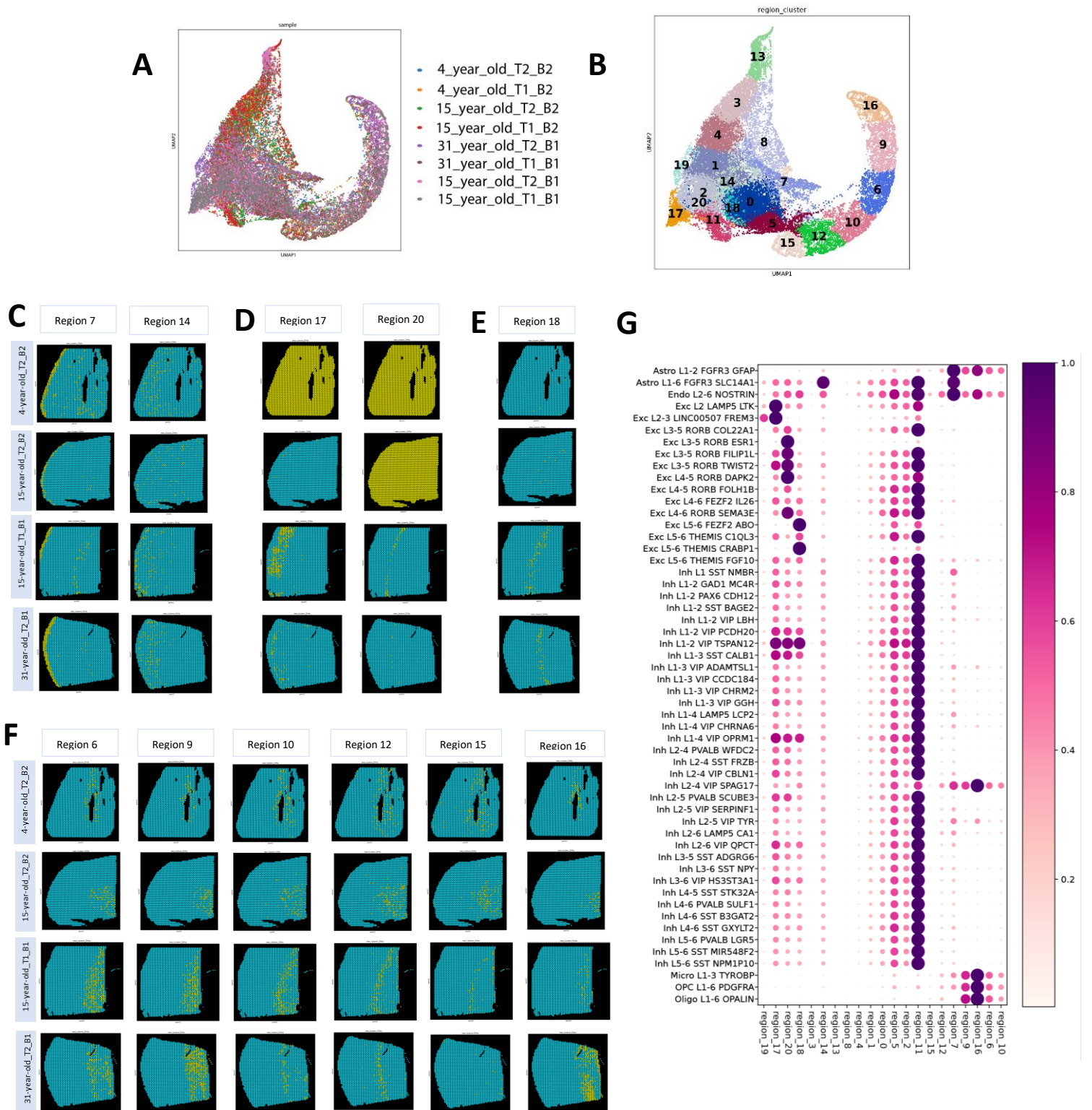


Figure 3.9: Identified tissue regions that differ in their cell composition.

A-B) UMAP plot of the merged Visium data coloured by sample and Leiden cluster respectively. C-F) Spatial plots showing the location of individual Leiden clusters in yellow in representative samples (note: yellow colour across the entire section in **D** indicates the absence of any spots for that cluster) (see github script `Annotating_clusters.py`). G) Dot plot with dot size and colour corresponding to the relative cell abundance of 53 cell types (rows) across regions clusters (columns).

Table 3.5: The total number of spots per Leiden cluster for each tissue samples in the 1st (B1) and 2nd (B2) Visium experiments.

Cluster	4_year_old_T1_ _B2	4_year_old_T2_ B2	15_year_old_T1_ B2	15_year_old_T2_ B2	15_year_old_T1_ B1	15_year_old_T2_ B1	31_year_old_T1_ B1	31_year_old_T2_ B1
0	59	170	127	337	556	469	876	494
1	372	423	402	347	252	335	240	309
2	164	180	292	228	703	598	197	196
3	593	519	477	394	7	19	34	18
4	408	421	373	374	28	40	41	160
5	32	133	98	306	165	159	621	323
6	101	80	28	89	317	267	265	200
7	203	326	122	76	120	124	150	216
8	214	379	284	260	25	20	101	50
9	76	62	25	77	247	217	121	398
10	130	117	54	93	249	219	243	97
11	192	234	318	279	20	70	17	9
12	145	140	96	145	177	99	233	103
13	131	120	270	75	12	222	246	13
14	104	176	57	67	159	286	12	108
15	173	154	130	105	69	40	96	16
16	69	69	6	43	60	121	53	361
17	0	0	19	8	335	240	8	69
18	1	8	8	9	92	56	20	55
19	1	0	0	2	8	4	0	174
20	0	0	0	0	50	18	0	7

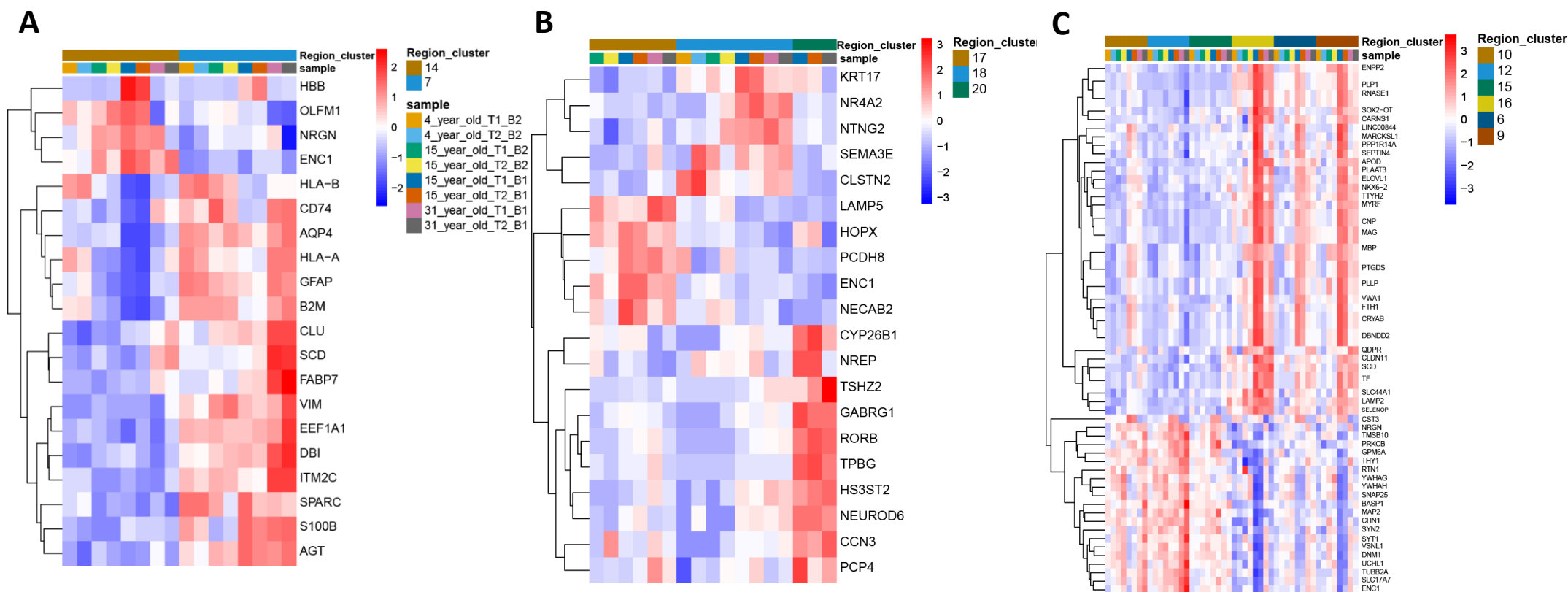


Figure 3.10: Heatmaps showing the top 20 differentially expressed genes associated with Leiden clusters of interest.

These include A) Region cluster 7 and 14, B) Region cluster 17, 18 and 20 and C) WM region clusters.

3.9 Optimization of in situ Hybridization chain reaction (HCR) using positive control probes

HCR was used to assess the gene expression profiles of several of the genes identified in the Visium data analysis. HCR is a technique which is used to detect specific mRNAs in tissue sections at a high resolution. Here, several mRNAs were detected at once using HCR. Custom DNA probes were ordered (see Table 2.18), each containing a sequence which enables hybridization to a specific mRNA molecule and a sequence that acts as an initiator of the HCR process i.e. the self-assembly of tethered fluorescent DNA hairpin polymers to amplify the gene expression signal.

The HCR protocol used in this study was adapted from (Kim et al., 2022). In their study they assessed the expression of excitatory neuron marker, *SLC17A7*, which was expressed by a subset of glutamatergic neurons and the inhibitory neuron marker, *GAD 1*, which was expressed by all inhibitory neurons. *AQP4* is an astrocyte marker known to be highly expressed in layer 1 (Hodge et al., 2019). In this study these probes were used as positive controls to set up the HCR protocol. Parameters such as fluorophore choice and probe concentration were assessed using the 46-year-old temporal cortex tissue from the Hockman lab biobank.

The effect of hairpin fluorophore on the probe signal was assessed using the *SLC17A7* probe (Fig. 3.11). Three different hairpins (Alexa 488, Alexa 514 & Alexa 594) were tested with each hairpin on a separate slide. The strongest signal was seen using the Alexa 594 and Alexa 488 hairpins while the Alexa 514 hairpin resulted in signal, however it was faint. In addition, autofluorescence from lipofuscin can be seen in the 594 channel, which shows clearly when using the 514 and 488 hairpins alone (Fig. 3.11). This assessment was also done for *GAD 1* using Alexa 594, 514, 647 and the strongest signal was seen in the Alexa 647 channel (Results not shown).

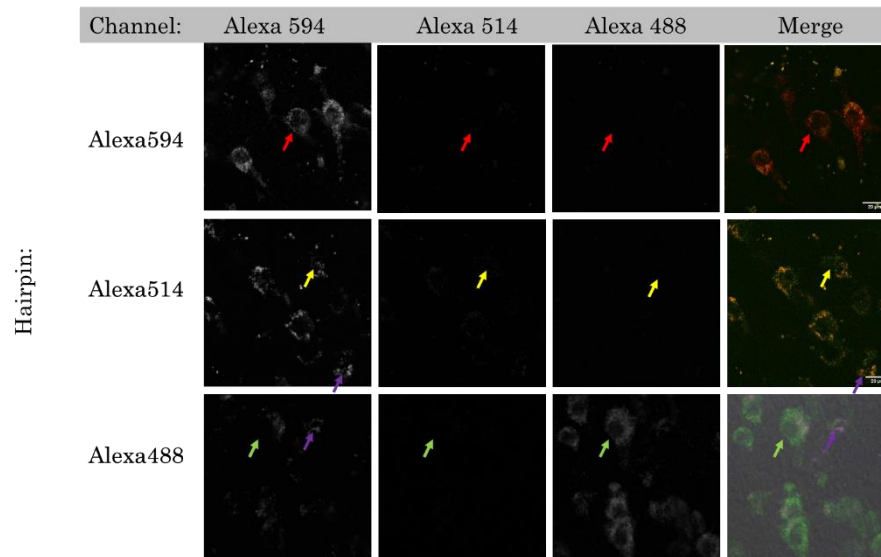


Figure 3. 11: Assessing the effect of hairpin fluorophore on the *SLC17A7* probe signal on 46-year-old tissue.
Arrows indicate a positive signal in the expected channel. Purple arrows indicate lipofuscin.

The effect of the probe concentration on the signal was tested using the positive control probes with the optimised hairpin fluorophores. Three different probe concentrations were tested (0.8, 1.6, 2.4 pmol) (Fig. 3.12). The probe concentration that gave the clearest signal for both *GAD1* and *SLC17A7* was 0.8 pmol. Lipofuscin was seen in all the different concentrations.

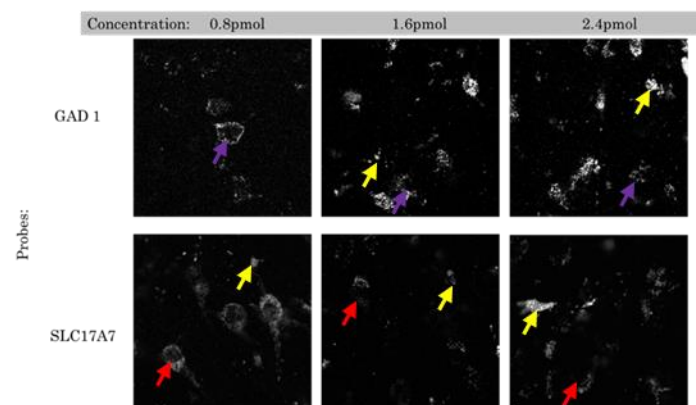


Figure 3.12: Assessing the effect of probe concentration on *Gad1* and *Slc17a7* probe signal on the 46-year-old.

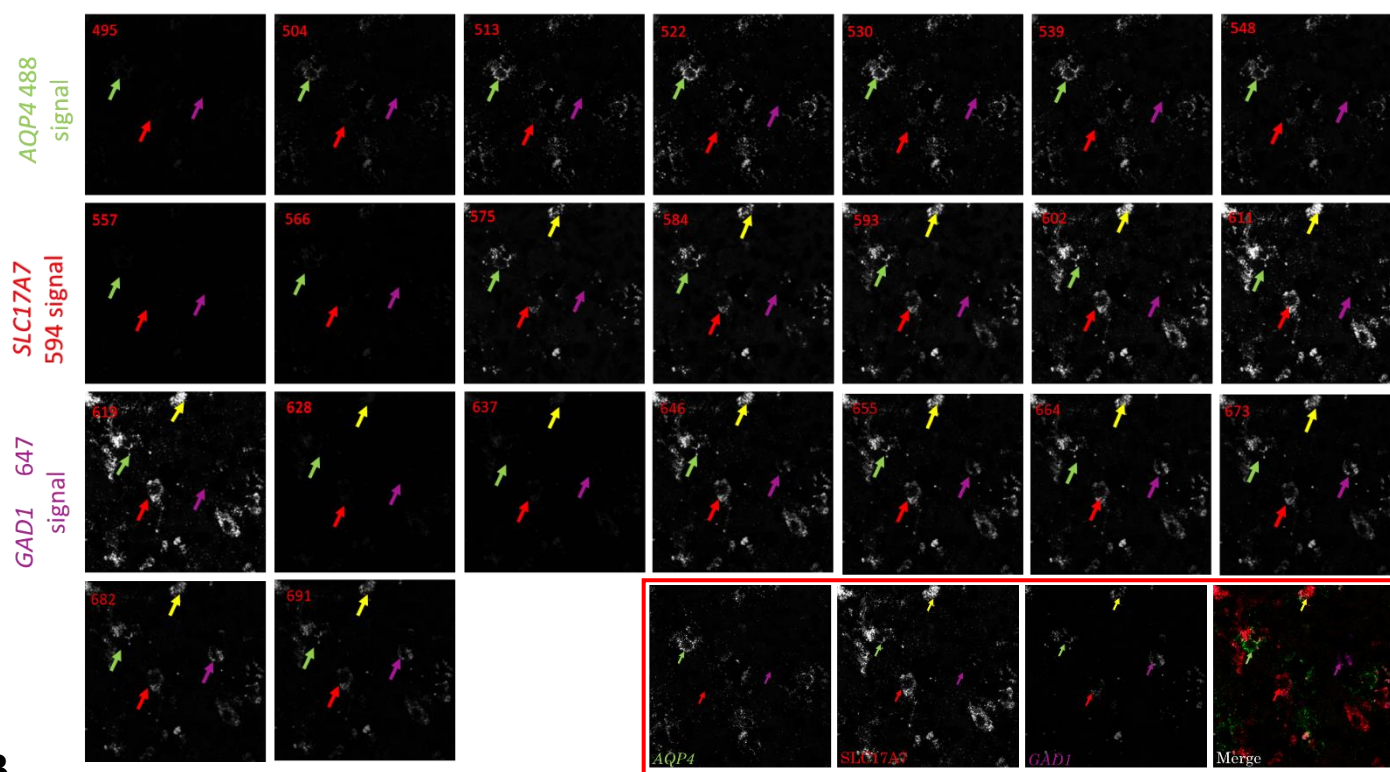
Red arrow = *SLC17A7*, Purple arrow = *GAD1*, Yellow arrow = Lipofuscin

Autofluorescence from lipofuscin is a common feature of human brain tissue (Moreno-García et al., 2018) and can complicate analysis and obscure mRNA signal. To overcome this, a technique called spectral imaging together with image analysis using linear unmixing can be used to separate mixed fluorescent signals and clearly show the contribution of each fluorophore, exposing autofluorescence signal that is present across many wavelengths. The first step in linear unmixing is to perform a lambda stack, which allows for the measurement of the emission signal at different wavelengths with essentially one spectrum indicated in each pixel (Fig. 3.13 A- B). Each fluorophore, regardless of the degree of spectral overlap with other probes, has a unique spectral signature or emission fingerprint that can be determined independently (Dickinson & Davidson, n.d.). The spectral signature is used to assign the real contribution of the fluorophore to each pixel in a lambda stack (Dickinson & Davidson, n.d.). In this study 3 different probes (*GAD1*: 647, *AQP4*: 488 and *SLC17A7*: 594) were tested together on one slide.

As can be clearly seen, the *AQP4* probe is positive in the 488 nm spectral range with the emission peak being 522 nm, after which the signal starts disappearing (Fig. 3.13 A-B; green arrows in row1). The *SLC17A7* probe is positive in the 594 nm spectral range with the emission peak at 619 nm (Fig. 3.13 A-B; red arrows in rows 2-3). The *GAD 1* signal is only seen in the 647 nm spectral range. The emission peak is 671 which can be seen with the brightest signal (Fig. 3.13 A-B; purple arrow in rows 3-4). Lipofuscin was mostly in the 594 nm spectral range (Fig. 3.13 A-B; yellow arrows). Moreover, the 594 nm signal bleeds through to the 647 nm channel. The Lambda stack provides the spectral data required for linear unmixing. Linear unmixing enables the precise determination of spectral profiles at every pixel in the image and therefore can reassign colour to regions that would otherwise appear mixed. This is seen in the outlined high magnification images (Fig. 3.13 A-B) in which a positive signal for each probe is detected without an overlap with other probes.

A

15-year-old



B

31-year-old

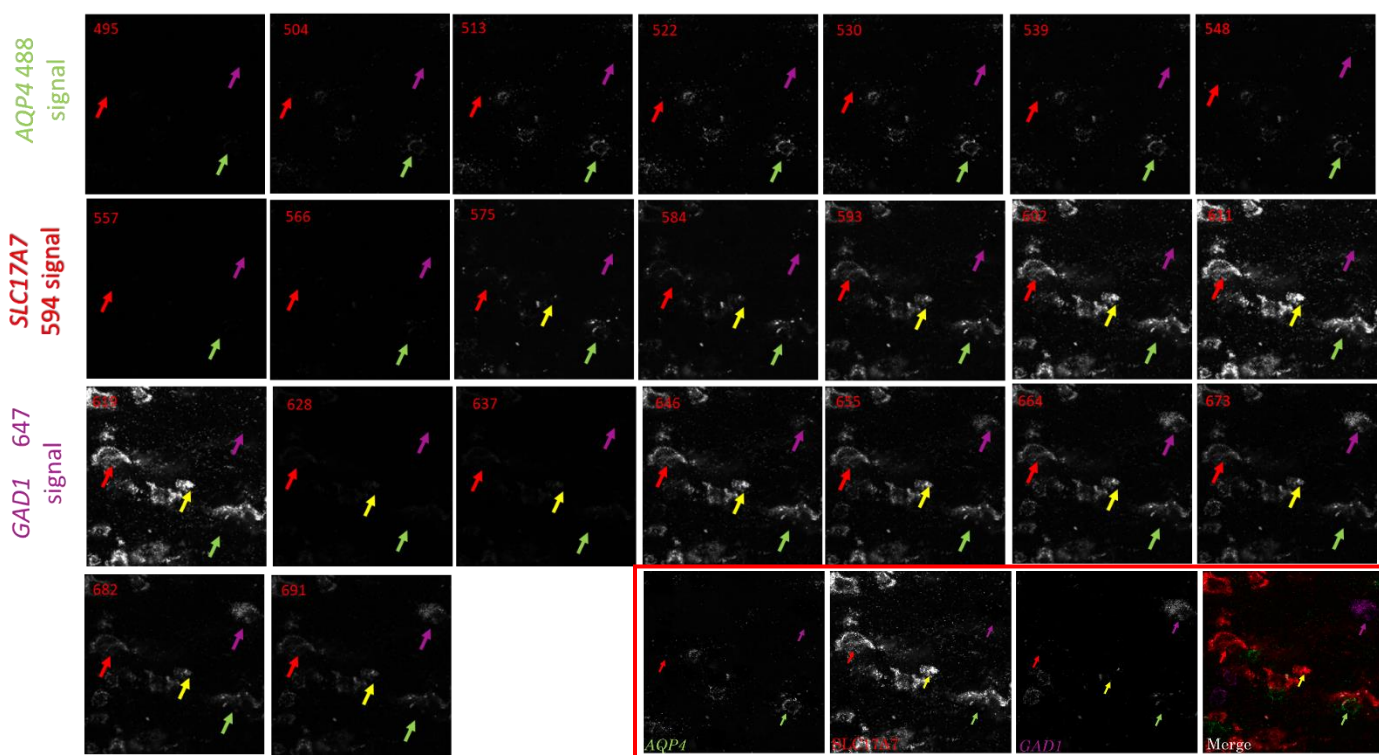


Figure 3.13: Multiplex in situ HCR using positive control probes on the human temporal cortex.

A) Lambda stack for the 15-year-old tissue and the corresponding view of each gene after spectral unmixing outlined in red. B) Lambda stack of the 31-year-old tissue and the corresponding view of each gene after spectral unmixing outlined in red. Arrows indicate regions of distinct signal for *AQP4* (green arrows), *SLC17A7* (red arrows), *GAD1* (purple arrows), and Lipofuscin (yellow arrows)

Tile scanning was used to capture the HCR signal for the positive control probes across the cortex layers (Fig. 3.14 A -B). *SLC17A7* and *GAD 1* were mutually exclusive in excitatory and inhibitory neurons, respectively. *AQP4* was highly expressed in astrocytes in layer 1, as well as scattered throughout the cortex (Fig. 3.14 A-B).

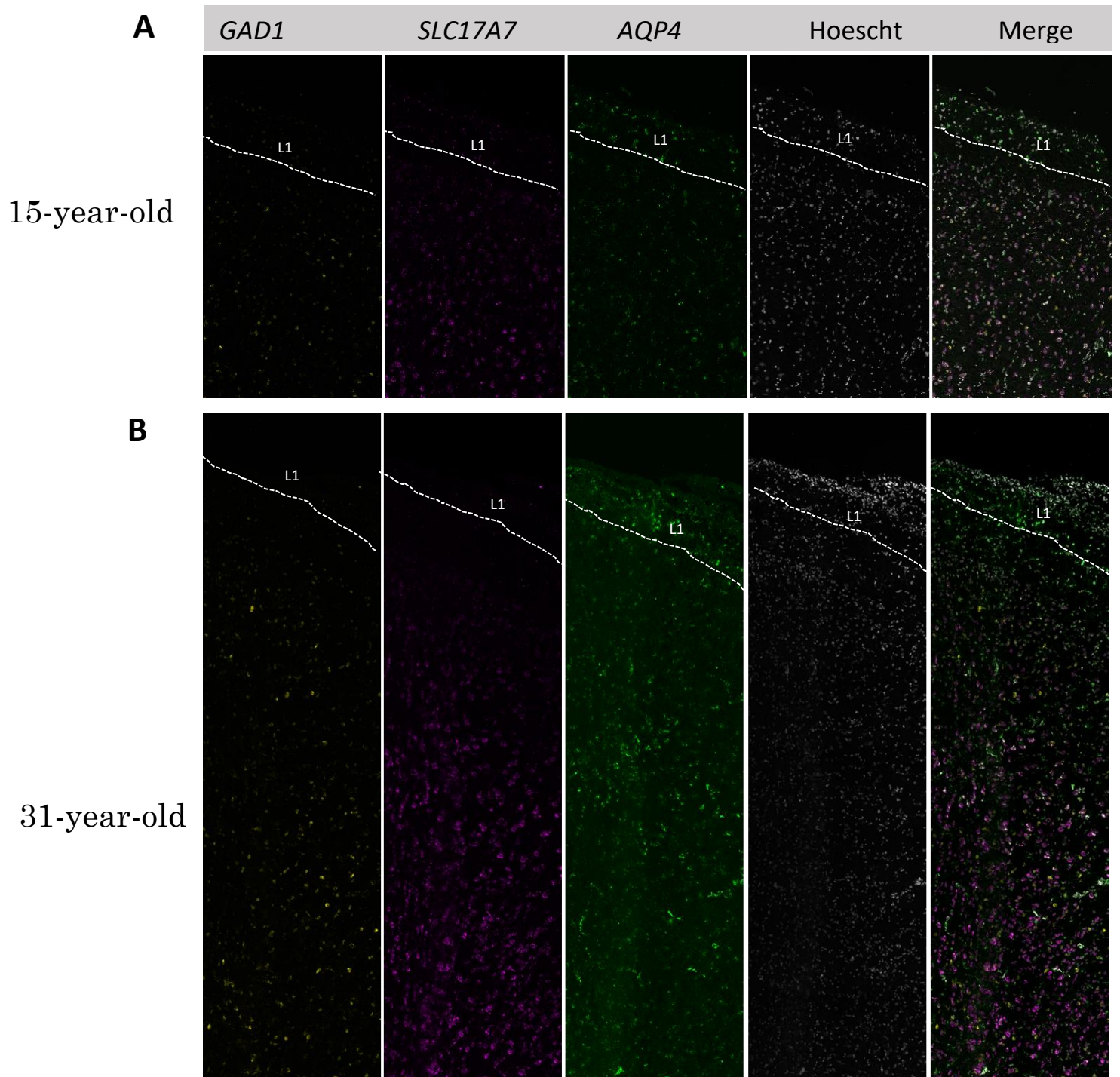


Figure 3.14: HCR Tile scan of the temporal cortex.

(A) 15-year-old tissue and (B) the 31-year-old tissue showing the expression of *GAD1*, *SLC17A7* and *AQP4*. Nuclei are stained with Hoescht. Dotted line demarcates the border of L1 determined by the high level of *AQP4* signal in this region.

3.10 Validating the Visium analysis using Bayesspace and HCR

Once the HCR had been optimised, the expression patterns of a selection of the genes identified in the *FindAllmarkers* analysis of the Visium datasets (see section 2.5.4), as well as selected known layer 1 markers, were further analysed using HCR and Bayesspace. Bayesspace is a statistical method that uses information from spatial neighbourhoods to provide a high-resolution readout of the expression profile of individual genes in spatial transcriptomic data (Zhao et al., 2021).

Layer 1 markers that were identified previously by Wu et al., and Maynard et al., in adult brain tissue were also analysed in this study (Maynard et al., 2021; Wu et al., 2022). Wu et al., previously showed *CXCL14* to be an upper layer marker of inhibitory neurons in the human visual cortex (Wu et al., 2022). Here, Bayesspace analysis of the Visium datasets confirmed that *CXCL14* is more highly expressed in cortical layer 1 in the temporal cortex than in the rest of the layers (Fig. 3.15 A). It can also be seen that the expression level appears to increase with age (Fig. 3.15 A). *AQP4* & *GAD1* Bayesspace expression profiles are shown as positive controls. To further analyse the *CXCL14* expression pattern, HCR was performed using the optimized *GAD1* and *AQP4* probes as positive controls. The HCR was performed on the same samples that were used in the Visium analysis. *GAD 1* marks all the inhibitory neurons and *AQP4* marks the astrocytes as well as cortical layer 1. The HCR analysis shows that *CXCL14* is highly expressed in *GAD1*-positive layer 1 inhibitory neurons across all the different ages (Fig. 3.15 B-C).

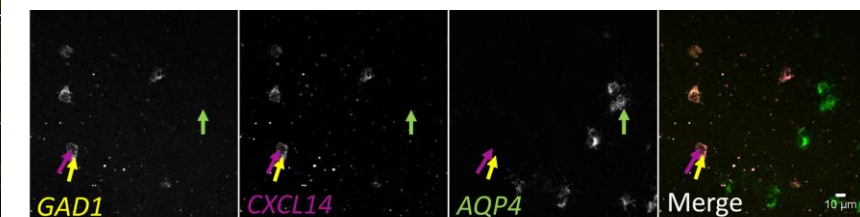
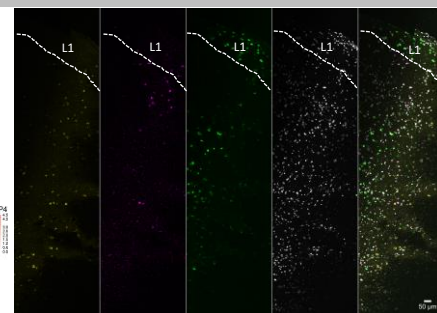
GAD1 CXCL14 AQP4 Hoescht Merge

A

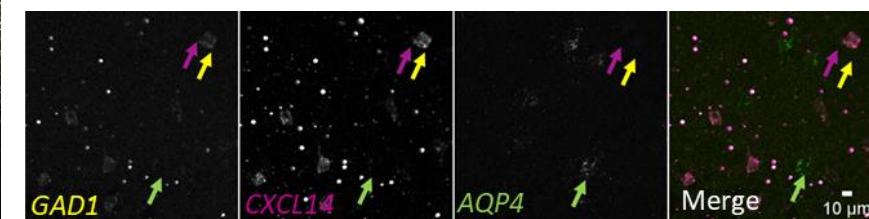
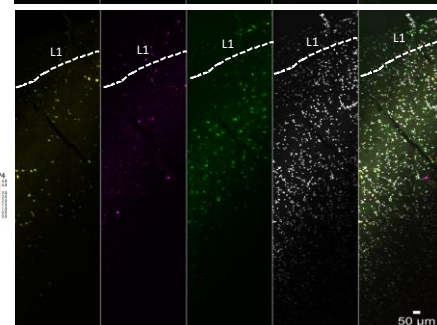
B

C

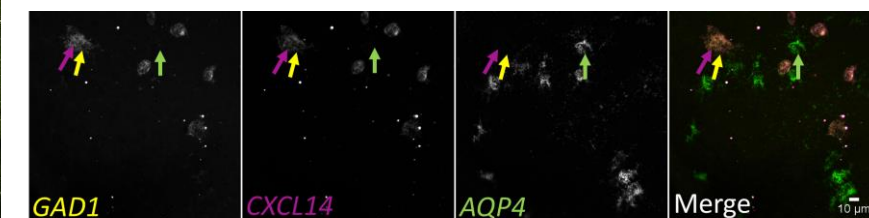
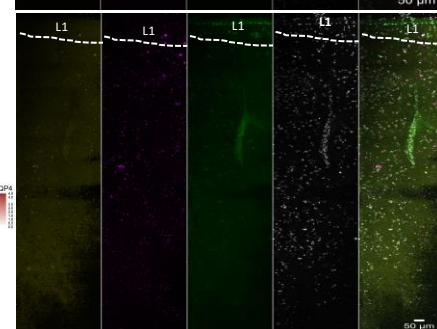
4-year-old_T1_B2



15-year-old_T2_B2



15-year-old_T1_B1



31-year-old_T2_B1

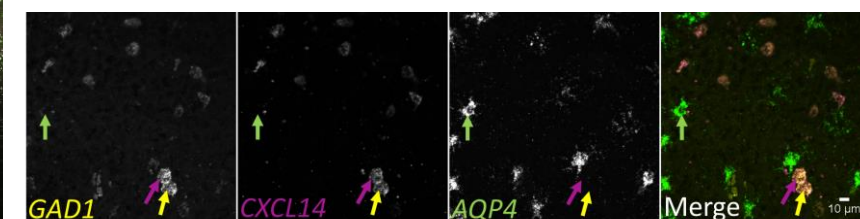
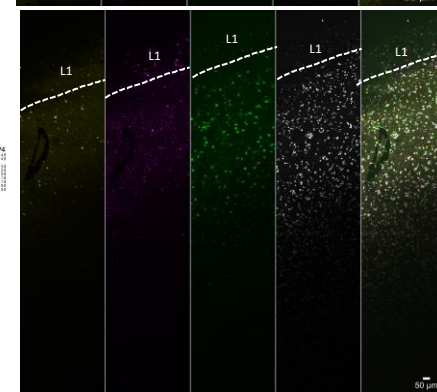


Figure 3.15: Bayesspace and multiplex *in situ* HCR analysis of layer 1-enriched gene, *CXCL14*, in temporal cortex tissue sections.

A) Enhanced resolution analysis of the Visium datasets using Bayesspace for *GAD1*, *CXCL14* and *AQP4*. B) Tile scans of tissue sections that were analysed with multiplex HCR showing the expression of *GAD1*, *CXCL14* and *AQP4*, counterstained with Hoescht (nuclei). C) High magnification views of the tile scans in B. Arrows indicate selected cells positive for *GAD1* (yellow), *CXCL14* (magenta) and *AQP4* (green) expression. L1, cortical layer 1 demarcated by a dotted line.

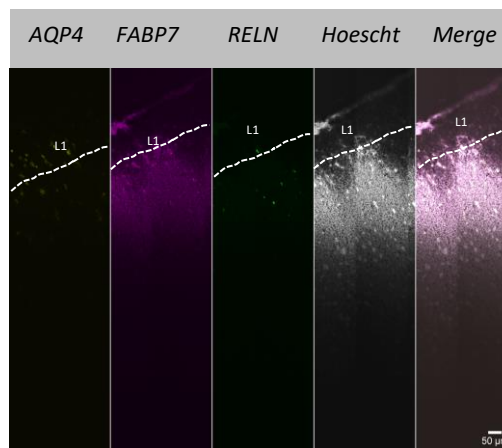
In their study, Maynard et al., confirmed a previously identified layer 1-enriched marker, *RELN*, and identified a novel layer 1-enriched marker gene, *FABP7*, in the adult human prefrontal cortex (Maynard et al., 2021). This study further confirmed *RELN* and *FABP7* to be highly enriched in layer 1 in the adult and paediatric temporal cortex samples. Using *FindAllmarkers*, *FABP7* was identified in the top 20 differentially expressed genes in the region cluster 7 of the 31-year-old sample only (see Fig. 3.10 C). *RELN* was not identified as a differentially expressed gene. Bayesspace analysis confirmed that *RELN* and *FABP7* are layer 1 markers across all the samples (Fig. 3.16 A). Both *RELN* and *FABP7* appear to increase in expression level with age (Fig. 3.16 A). HCR analysis further validated the expression of *RELN* and *FABP7* in cortical layer 1 (Fig. 3.16 B-C). *FABP7* is known to be expressed in astrocytes (Ebrahimi et al., 2015) and *RELN* is expressed by Cajal-Retzius cells (Vilchez-Acosta et al., 2022). As expected *FABP7* co-located with the positive control *AQP4* (Fig. 3.16 B-C).

A

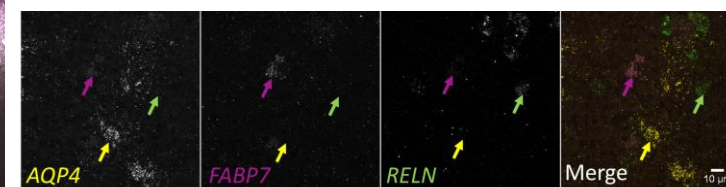
15-year-old_T2_B2



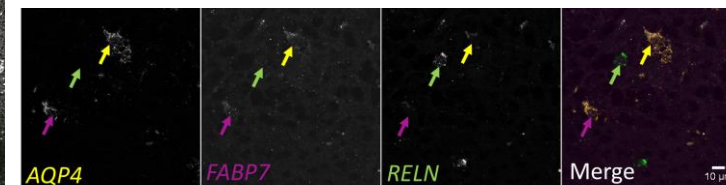
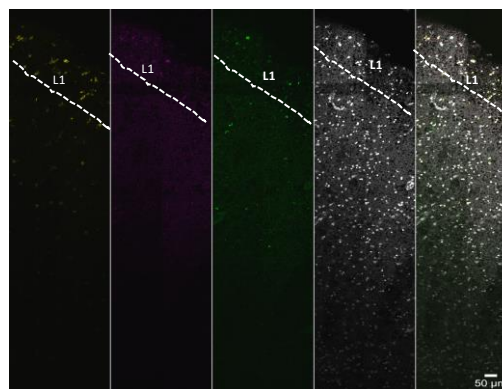
B



C



15-year-old_T1_B1



31-year-old_T2_B1

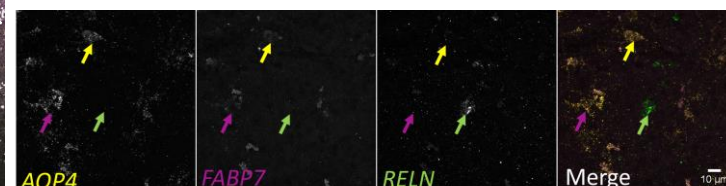
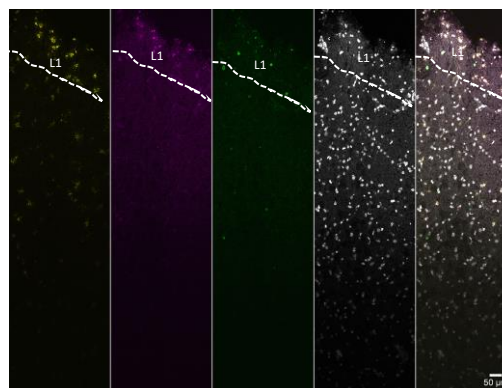


Figure 3.16: Bayesspace and multiplex *in situ* HCR analysis of layer 1-enriched genes *RELN*, *FABP7* and *AQP4*, in temporal cortex tissue sections.

A) Enhanced resolution analysis of the Visium datasets using BayesSpace for *AQP4*, *FABP7* and *RELN*. B) Tile scans of tissue sections that were analysed with multiplex HCR showing the expression of *AQP4*, *FABP7* and *RELN*, counterstained with Hoescht (nuclei). C) High magnification views of the tile scans in B. Arrows indicate selected cells positive for *AQP4* (yellow), *FABP7* (magenta) and *RELN* (green) expression. L1, cortical layer 1 demarcated by a dotted line.

TSHZ2 and *RORB* were identified in the top 20 of the differentially expressed genes of region cluster 20, which clearly marked layer 4 in the 1st Visium experiment (see Fig 3.10 B). Bayesspace analysis shows that *TSHZ2* is expressed in low amounts across all the samples and is more prominent expression in layer 4/5 in the 15-year-old and 31-year-old samples from the 1st Visium experiment, with higher expression in the adult tissue (Fig. 3.17 A). *RORB* is expressed in low amounts in the 4-year-old sample and 15-year-old samples from the 2nd Visium experiment compared to the 15-year-old and 31-year-old samples from the 1st Visium experiment where it is highly expressed in layer 4/5 (Fig. 3.17 A).

CLSTN2 was in the top 20 differentially expressed genes of region cluster 18 (see Fig 3.10 B), which clearly marked cortical layer 5/6 in both the 15-year-old and 31-year-old samples analysed in the 1st Visium experiment. Cluster 18 marked a few spots in cortical layer 5/6 in the 4-year-old and 15-year-old sampled analysed in the 2nd Visium experiment. *CLSTN2* was highly expressed in the 4-year-old samples compared to the rest of the samples. Bayesspace analysis revealed that *CLSTN2* shows a more prominent expression in cortical layers 5/6 in the 15-year-old and 31-year-old samples from the 1st Visium experiment compared to the 4-year-old sample and 15-year-old samples from the 2nd Visium experiment, where it is expressed in low amounts, and it is not layer enriched (Fig. 3.17 A). It is important to note that the 2nd Visium datasets were not of good quality based on the QC analysis results.

In the HCR analysis, *RORB* is seen in the layer 4 in all the samples. *TSHZ2* is highly expressed in layer 4 and 5, but it is sparser in layer 5, as seen in the Bayesspace images. *CLSTN2* signal is present in all the cortical layers in the 4-year-old sample but does appear higher in the deeper layers corresponding to layers 5/6 (Fig. 3.17 B). In the two 15-year-old samples and the 31-year-old samples, *CLSTN2* is more clearly layer 5/6- enriched (Fig. 3.17 B). *RORB*, *CLSTN2* and *TSHZ2* co-locate in layer 4 and 5 only (see Fig. 3.17 C). Even though they co-locate, *CLSTN2* positive cells are sparser (Fig. 3.17 B).

The use of HCR has proven to be a valuable tool in the investigation of gene expression patterns. By detecting and amplifying specific RNA molecules within cells, this technique allows for the visualization and localization of gene expression with high specificity and sensitivity. These results demonstrated the utility of HCR in validating the Visium experiment and detecting genes that co-locate and those that are layer-enriched.

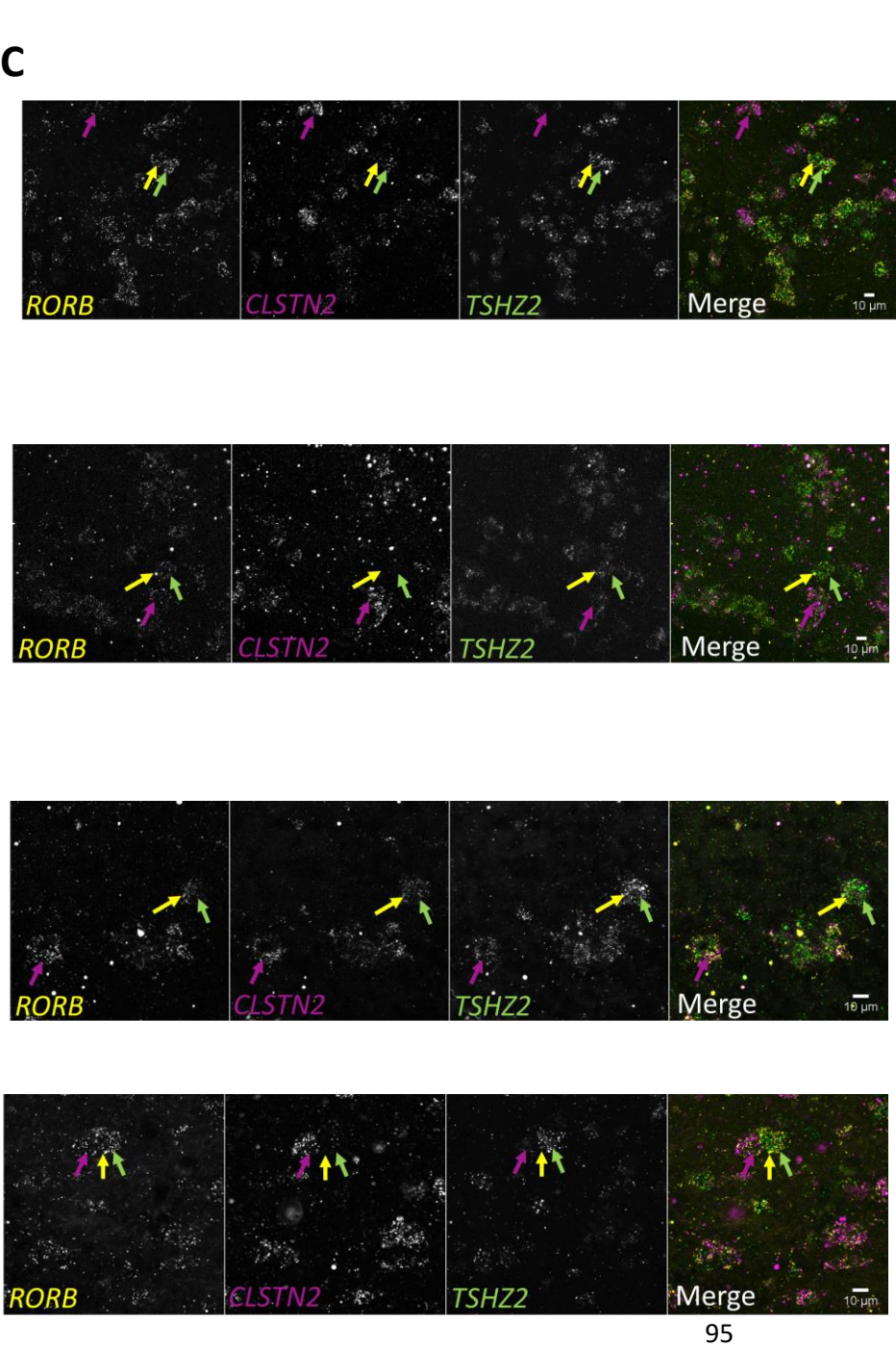
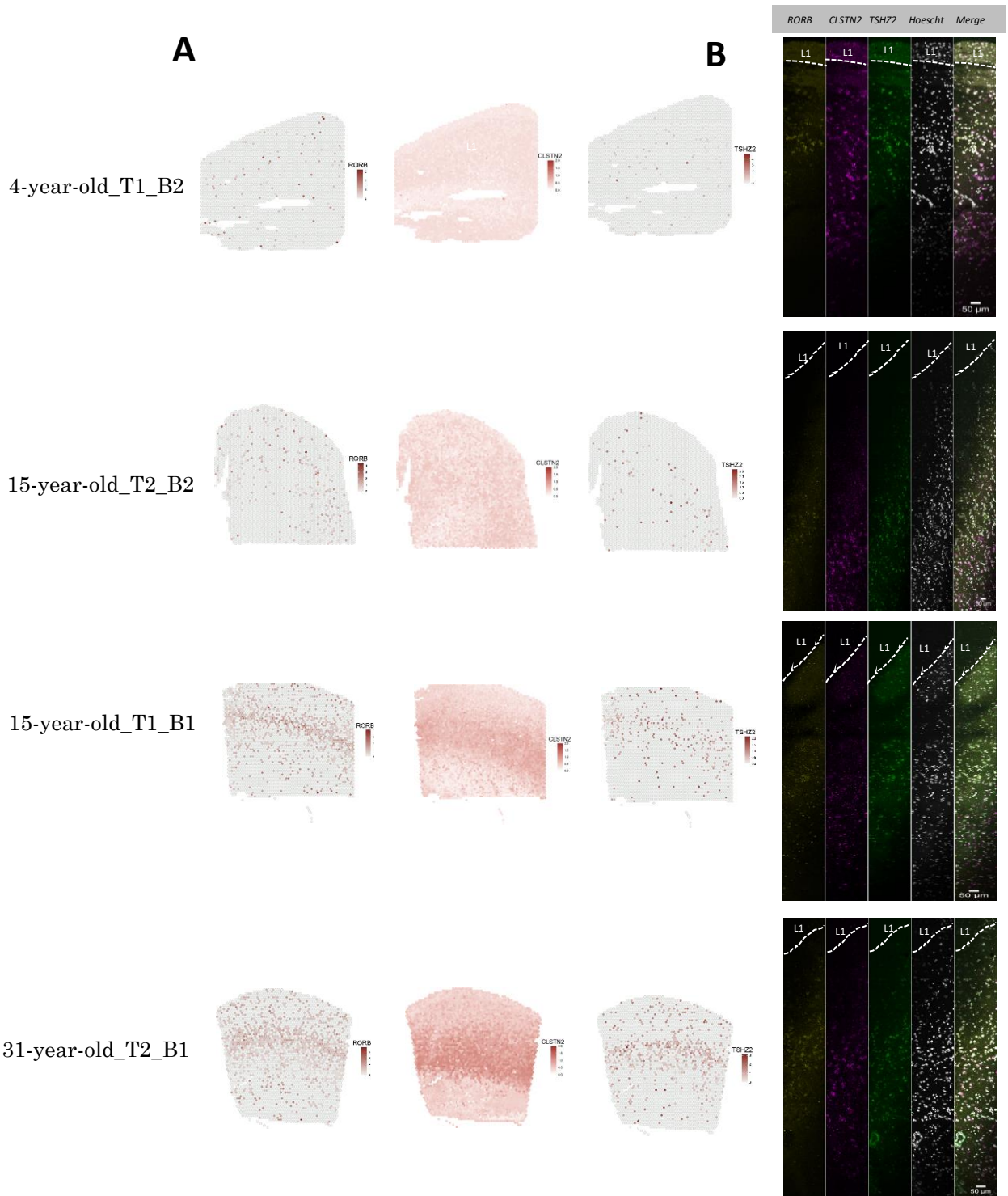


Figure 3. 17: Bayesspace and multiplex in situ HCR analysis of layer 4-6-enriched genes, *RORB*, *CLSTN2* and *TSHZ2*, in temporal cortex tissue sections.

A) Enhanced resolution analysis of the Visium datasets using BayesSpace for *RORB*, *CLSTN2*, *TSHZ2*. B) Tile scans of tissue sections that were analysed with multiplex HCR showing the expression of *RORB*, *CLSTN2* and *TSHZ2*, counterstained with Hoescht (nuclei). C) High magnification views of the tile scans in B. Arrows indicate selected cells positive for *RORB* (yellow), *CLSTN2* (magenta) and *TSHZ2* (green) expression in L4-6. L1, cortical layer 1 demarcated by a dotted line.

CHAPTER 4: DISCUSSION

The aim of this study was to contribute to the cell atlas of the maturing human brain by analysing spatial gene expression profiles in bio-banked ante-mortem temporal cortex samples over the course of brain maturation using the Visium technology from 10X Genomics. While Maynard et al., utilized this technology on an adult human dorsolateral prefrontal cortex (Maynard et al., 2021), the present study focuses on paediatric samples to facilitate comparisons of human brain maturation. Moreover, this study used a variety of different bioinformatics methods including cell2location and BayesSpace to analyse the spatial transcriptomic data to determine the spatial locations of cell types and explore how spatial gene expression patterns change over time. Additionally, HCR was used to validate the spatial gene expression patterns seen in the cell2location and BayesSpace results.

4.1 Tissue preservation

Prior to performing the Visium experiment it was important to ensure that the tissue was well preserved to limit the RNA degradation which can affect the transcriptomic data. There are several factors that causes RNA degradation such as in vivo during surgery, ex vivo during the transportation to the lab and during thawing and manipulation after embedding (Somiari et al., 2022). For this study it was important to prepare and store the tissue in a way that preserves RNA integrity, spatial orientation, and morphology of the tissue. To achieve this, all the samples were embedded in OCT and frozen in liquid nitrogen prior to storing them at -80°C. Embedding tissue in OCT preserves the morphology of the tissue and helps in maintaining an optimal temperature during sectioning. Formalin fixation and paraffin embedding is another method for preserving tissue, it can preserve the morphology and cellular details of a tissue (Kokkat et al., 2013). Visium is also capable of profiling FFPE tissues, however this technique suffers from false signal caused by long-noncoding RNAs (lncRNAs) due to their low expression levels and potential for cross-hybridization with other transcripts (Prakrithi et al 2023).

4.2 Quality control of Visium libraries

Quality checking of the Visium libraries at the cDNA amplification stage and post-PCR indexing QC stage were important steps to ensure good library preparation and a generation of good quality sequencing data. The cDNA amplification plot shows the amount of cDNA and

distribution of fragment sizes in the sample library after amplification. Enough cDNA for subsequent steps is always available regardless of a degraded RNA input. The fragment size distribution influences the library construction and sequencing results therefore, it is important to ensure that the fragment size distribution is within the optimal range. In this study, the fragments size distribution of the samples was within or above the recommended range (for the post cDNA amplification the recommended range is 350- 1000bp with a peak around 600-900bp and for the final library the range is 300-1000bp and a peak at 423-478bp). According to Antico et al., this is an indication of good quality input RNA (Antico et al., 2022). Even though the optimal fragment size distribution of the 2nd Visium experiment was within the recommended range, the sequencing metrics, such as the Median UMI count and median genes per spot, were lower than the 1st Visium experiment. According to 10X Genomics low UMI values could be biological or maybe be because of a low sequencing depth, library complexity and sample quality. In this study, the low UMI values could have been due to lower quality samples. This might have been caused by technical problems during sectioning such as fluctuation in the temperature of the cryostat, which could have led to RNA degradation. Interestingly, the total number of genes detected appeared to not be affected by the lower UMI per spot readings. To further investigate and troubleshoot, it will be important for future studies to ensure the correct cryostat temperature during the tissue sample capture step to ensure good RNA quality. This can be done by performing controlled experiments where cryosectioning is conducted under stable and well-controlled temperature conditions. This would be beneficial in confirming the impact of temperature fluctuations on section quality.

4.3 Spatially mapping cell types using cell2location.

scRNA-seq methods can provide gene expression information at a single cell level however these methods lose the spatial context of the cells. Visium spatial gene expression provides the spatial information of cell types, but each spot covers multiple cells. According to 10X Genomics each spot can capture 1-10 cells. However, this study demonstrated that some spots could capture more than 10 cells, which makes it difficult to assign each spot to a single cell type. To spatially map the developing human brain cell types *in situ*, a deconvolution method called cell2location was used which aims to identify cell types and their relative proportions contributing to a spot. It takes advantage of the complementary strengths of sc/snRNAseq and Visium datasets, allowing for integration of these two data types

(Kleshchevnikov et al., 2022). Here, cell2location was used to integrate the Visium mRNA count matrices with existing temporal cortex snRNA-seq data from the Hockman lab that had been annotated using the published adult middle temporal gyrus brain map dataset (Hodge et al., 2019). Using this method 54 cell sub-types were identified across all the samples that represent Inhibitory neurons (34 sub-types), excitatory neurons (14 sub-types) and non-neuronal cell types (6 sub-types). Oligodendrocytes subtype *OLIGO L1-6 OPALIN* was the most abundant cell type, which is consistent with previous spatial transcriptomic analyses of the human temporal cortex (Fang et al., 2022; Langseth et al., 2021). They identified higher abundances of non-neuronal cell types than neuronal cell types, with oligodendrocytes being the most abundant in Fang et al. (Fang et al., 2022).

The astrocyte subtype *Astro L1-2 FRGF3 GFAP* was highly enriched in layer 1 and the WM across all the samples. In contrast, Hodge et al., and Langseth et al., who used microdissection of cortical layers followed by scRNA-seq and pci-Seq respectively, only describe the distribution of this subtype in the cortical layers (Hodge et al., 2019; Langseth et al., 2021). Hodge et al., shows the majority of *Astro L1-2 FRGF3 GFAP* cells in layer 1, with a few cells in layer 6 (Hodge et al., 2019). Langseth et al. shows majority of this subtype in layer 3 (Langseth et al., 2021). More recently, Fang et al. used MERFISH to describe the presence of an astrocyte subtype (*ASCxi*) that is enriched in layer 1 and the WM (Fang et al., 2022). This astrocyte subtype could likely correspond to the astrocytes seen in this study. Additionally, in this study, oligodendrocytes were highly enriched in the WM just as been reported by previous studies (Fang et al., 2022; Hodge et al., 2019; Langseth et al., 2021).

Previous spatial transcriptomic studies also showed that excitatory neurons form layer-specific spatial patterns in the adult human brain. Tasic et al. also showed that excitatory neurons span cortical layers in the adult mouse brain (Tasic et al., 2018). In this study most excitatory neurons were in the expected cortical layers in the adult sample. Interestingly this was also seen in the juvenile sample from the 1st Visium experiment (15-year-old_T2_B1) which has not been previously reported. Previous reports have shown that *Exc L2 LAMP5 LTK* is less layer restricted than expected (Hodge et al., 2019). In correspondence, this study found *Exc L2 LAMP5 LTK* to be highly enriched in layer 2 and layer 3. Hodge et al., found a few cells of this cell type in layer 1 (Hodge et al., 2019), however this was not seen in this study. In Hodge et al.'s data there was no evidence of this cell type in layer 6. However, this study

showed evidence of *Exc L2 LAMP5 LTK* in L6 in all the samples corresponding to what was previously reported by Langseth et al., 2021. To further confirm this distribution, future studies could use HCR to analyse the distribution of the marker genes for this cell type. The samples from the 2nd Visium experiment did not show layer-specific excitatory neurons, however this could be because of the lower quality of these samples. In contrast to excitatory neurons, inhibitory neurons are not as layer specific as the excitatory neurons, as previously suggested by Tasic et al. for the adult mouse brain, but rather were often shared across many cortical layers (Tasic et al., 2018). In support of this finding in the current study, *Inh L1-4 LAMP5 LCP2* was enriched in layer 1 but present sparsely in all the cortical layers, which is also consistent with what Hodge et al. 2019 found.

4.4 Understanding the tissue organization using cell clustering.

Cell2location was used to identify spatial occurrence of cell types. To better understand tissue organization and compartmentalisation, NMF was performed using the cell abundance profiles identified by cell2location (Kleshchevnikov et al., 2022). The NMF decomposition grouped the spatial cell type abundance profiles into compartments that capture co-localized and distinct cell types. Multiple distinct factors were extracted with well-defined spatial patterns. Another method was conducted called Leiden clustering to identify tissue regions. NMF groups cell types that may be interacting into 1 factor. It also considers that multiple cell types can co-exist at the same Visium location (Kleshchevnikov et al., 2022). Leiden clustering reveals non-overlapping tissue regions defined by cell types. These two methods can give an understanding of the tissue organization by giving insight into the overall tissue architecture (Kleshchevnikov et al., 2022)

4.4.1 White matter clusters

In both the NMF and Leiden clustering analysis, the WM consisted of oligodendrocytes, microglia, astrocytes, endothelial and OPCs. Previous spatial transcriptomic analyses showed a high abundance of *Oligo L1-6 OPALIN* in layer 6 and the WM (Langseth et al., 2021), however this study showed that this cell type has a distinct distribution, with a high abundance level in the WM only, as shown by the distribution of NMF factor 12. The lack of *Oligo L1-6 OPALIN* in the cortex could be due to the Visium technique which does not result in as high resolution data as pci-Seq (Langseth et al., 2021) and therefore was not able to pick up the oligodendrocyte signal in layer 6. Leiden cluster analysis showed that the WM also includes

inhibitory neurons, such as *Inh L2-4 VIP SPAG17*, which was widely spread across the tissue and highly enriched in the WM. This finding expands on Hodge et al., and Langseth et al., findings that this cell type was found in layers 1-4 (Hodge et al., 2019; Langseth et al., 2021). Inhibitory neurons found in the white matter of the brain are responsible for regulating the transmission of signals between different areas of the brain. They release inhibitory neurotransmitters, such as GABA, which can dampen the activity of other neurons (Sedmak and Judaš, 2021). These inhibitory neurons play a critical role in maintaining the balance of excitatory and inhibitory signals in the brain, and an imbalance in these signals can lead to neurological disorders (Scaduto et al., 2022).

Highly expressed genes in the six WM clusters that were identified by Leiden clustering included *MBP*, *CST3* and *SOX2-OT*. Myelin basic protein (MBP) is a gene that codes for an essential protein that is involved in the formation of myelin sheaths. It is a well-known WM marker (Maynard et al., 2021) expressed by oligodendrocytes. Not only is *MBP* highly expressed in the white matter of human brain but it was previously reported to be highly expressed in the white matter of an adult rat brain (Sampaio-Baptista et al., 2020). *Cystatin C* (*CST3*) has been reported to be highly expressed in astrocytes in the mouse brain across different ages (Farhy-Tselnicker et al., 2021). It has been associated with Alzheimer disease where it functions as a protective factor by inhibiting cysteine proteases (Pan et al., 2020). Together with this study these findings suggests that *CST3* is primarily expressed in astrocytes. This means that *CST3* could be a useful marker for identifying and studying astrocytes in the brain. *SOX2-OT* is a long noncoding RNA involved in neural differentiation (Li et al., 2020) including inhibitory interneurons (Li et al., 2020) and plays a role in gene expression during development (Tosetti et al., 2017). Dysregulation of *SOX2-OT* has been implicated in the pathogenesis of psychiatric disorders such as schizophrenia and bipolar disorder, which have been associated with altered inhibitory neurotransmission (Li et al., 2020). This study identified *SOX2-OT* in the WM which may suggest that it could be a useful marker for inhibitory neurons found in the WM.

4.4.2 Astrocyte distribution and associated layer 1 marker, *FABP7*

Two astrocytes' subtypes were identified in this study, and they had different laminar distributions: *Astro L1-6 FGFR3 SLC14A1* was seen in layers 1-6 and *Astro L1-2 FGFR3 GFAP* was highly enriched in layer 1 in all the samples. Leiden cluster 7 contained both astrocytes'

subtypes (*Astro L1-2 FGFR3 GFAP* in high abundance and *Astro L1-6 FGFR3 SLC14A1* in low abundance) and it marked layers 1 and 6. This validated what was previously reported about the spatial location of astrocytes in the adult human brain by Hodge et al., 2019. In the NMF analysis *Astro L1-2 FGFR3 GFAP* was distinctly associated with factor 9, which was highly enriched in layer 1 and found in the WM in all the samples. This group possibly corresponds to the ASCxi sub-type described by Fang et al., which showed a similar distribution (Fang et al., 2022). Interestingly, factor 9 and the corresponding astrocyte sub-type showed a decrease in abundance in the WM with age. This finding could be linked to the role that astrocytes play in regulating the production and turnover of myelin, the fatty insulation that surrounds axons in the white matter (Bugiani et al., 2022) and their importance in aging-related changes in the brain (Verkhatsky et al., 2021). Therefore, a decrease in the number or function of astrocytes in the white matter could therefore lead to structural and functional changes in the white matter, including demyelination, disruption of ion homeostasis, and impaired synaptic function. This may be a key contributor to age-related cognitive decline and neurodegenerative diseases. *Astro L1-6 FGFR3 SLC14A1* was associated with NMF factor 8 and Leiden cluster 14, which was highly enriched in layer 1 but also found across all the layers just as reported by Hodge et al., 2019.

Differential expression analysis was performed for Leiden cluster 7 and 14 to identify genes that were differentially expressed in these two regions. Well known L1 genes were expressed in cluster 7 such as *AQP4* and *GFAP* (Hodge et al., 2019; Maynard et al., 2021). This finding was validated by the Bayesspace and HCR analysis of *AQP4*, which showed expression of *AQP4* in layer 1 in all samples. Furthermore, these analyses confirmed a previously identified layer 1 marker, *FABP7* (Maynard et al., 2021), which was differentially expressed in region cluster 7 and showed co-localisation with astrocyte marker *AQP4* in layer 1. *FABP7* is known to be enriched in A1-like astrocytes and it is important for cell growth and differentiation. Clinically, *FABP7* has been implicated in neurodegenerative diseases such as Alzheimer's. *FABP7* may contribute to AD pathology by promoting the accumulation of beta-amyloid (which is a harmful protein that is commonly found in the brains of people with Alzheimer's disease (AD)) and neuroinflammation, while also being upregulated in the brains of AD patients and animal models of AD (Needham et al., 2022). Interestingly, the Bayesspace analysis for *FABP7* shows an increase in expression with age. This finding could be linked to its role in cell growth and

differentiation. According to Clarke et al., an increase in A1-like astrocytes with aging could play a role in exposing aging brains to neurodegenerative diseases (Clarke et al., 2018). This was shown in a study by Killoy et al., in which an upregulation of *FABP7* in a co-culture model of astrocytes and motor neurons promotes a pro-inflammatory response in astrocytes that is toxic to motor neurons, possibly enhancing the susceptibility of aging brains to neurodegeneration (Killoy et al., 2020).

4.4.3 Excitatory and inhibitory neuron distribution and co-localisation

All the *RORB* sub-type excitatory neurons were enriched in layers 3 to 6, which is consistent with previous spatial transcriptomic studies of the temporal cortex (Hodge et al., 2019; Langseth et al., 2021). In the NMF analysis, the L3-5 *RORB* sub-types colocalised (factor 2) and L4-6 *RORB* sub-types colocalised (factor 10). In the Leiden cluster analysis these excitatory neuron sub-types were highly enriched in region cluster 20, which distinctly marked cortical layer 4 in the samples from the 2nd Visium experiment. As expected, *RORB* was one of the top 20 differentially expressed genes associated with Leiden cluster 20. Interestingly, the distribution of *Exc L3-5 RORB ESR1* in the cortex shows a change with age. In the 4-year-old this cell type is spread across all cortical layers, and it becomes highly enriched in layer 4 with brain maturation. This could explain the lack of region cluster 20 in the 4-year-old sample. Similarly, Bayesspace analysis showed low, sparse *RORB* expression in the 4-year-old and 15-year-old samples from 2nd Visium experiment and enriched expression in layer 4/5 in the 15-year-old and 31-year-old samples from the 1st Visium experiment. In contrast, HCR analysis showed *RORB* expression in cortical layers 4/5 in all the samples. Therefore, the trend seen in the Visium and BayesSpace analysis needs to be further explored by adding more samples to this study, as it could also be a patient-specific phenotype. *RORB* is a gene that codes for Retinoic acid-related orphan receptor beta (*RORβ*), which is transcription factor and a marker of layer 4 neurons. Importantly, *RORB* function has been reported to be disrupted in epilepsy cases (Rudolf et al., 2016). This study used tissue samples that were from patients with epilepsy, so it is possible that the tissue samples used in the 2nd Visium experiment do not represent normal *RORB* expression.

Several scRNA-seq studies have reported *Exc L5-6 FEZF2 ABO* and *Exc L5-6 THEMIS CIQL3* to be highly enriched in layers 5/6 in the adult human brain (Hodge et al., 2019; Langseth et al., 2021). Consistent with these studies, *Exc L5-6 FEZF2 ABO* and *Exc L5-6 THEMIS CIQL3* were

spatially enriched in layers 5/6 of the paediatric and adult temporal cortex. This was seen in the two methods used to understand the tissue organization (Leiden clustering: cluster 18 and NMF: factor 14). In a scRNA-seq study conducted by Pfisterer et al., (2020) the subtype *L5-6 FEZF2 TLE4 ABO* was found to be associated with epilepsy (Pfisterer et al., 2020). It exhibited one of the largest transcriptomic differences between epileptic and nonepileptic cortices among the principal neuron subtypes analysed. The gene expression correlation analysis and enrichment analysis of differentially expressed genes confirmed the prevalence of epilepsy-associated genes within this subtype. Therefore, the findings suggest that the *L5-6 FEZF2 TLE4 ABO* subtype of principal neurons may play a role in epilepsy (Pfisterer et al., 2020). *CLSTN2* was identified as a gene associated with this region. According to Lipina et al., *CLSTN2* in the mouse brain is linked to cognitive performance, memory behaviour, and a decrease in the number of functional inhibitory synapses (Lipina et al., 2016). In adolescents, *CLSTN2* may play a role in mnemonic processing in and may be involved in the vulnerability to environmental exposures that affect cognitive function (Jacobsen et al., 2009). In the old age population, it was reported that *CLSTN2* may play a role in cognitive function, along with other genes involved in synaptic plasticity (Laukka et al., 2013). More recently, *CLSTN2* has been shown to be expressed in the deep layers of the cortex (Zhou et al., 2020). In the Bayesspace analysis, *CLSTN2* was enriched in L5/6 of the 1st Visium experiment samples. HCR validated the expression of *CLSTN2* in the deep layers of the cortex, but showed it is also expressed in other layers at lower levels.

TSHZ2 was associated with clusters 18 and 20, however it was highly enriched in cluster 20 of the adult sample, which marks layer 4. *TSHZ2* was previously identified as a marker of layer 5 excitatory neurons in the mouse primary cortex (Yao et al., 2021) and in the mouse neocortical areas (primary visual cortex and the anterior lateral motor cortex) (Tasic et al., 2018). *TSHZ2* is a transcription factor associated with establishing a gradient of gene expression that contributes to the formation of functional areas during cortical development (Yao et al., 2021). In the HCR analysis, *RORB* and *TSHZ2* were colocalised, with *RORB* being layer 4 specific and *TSHZ2* located in layers 4/5 in all the samples. Overall, HCR analyses of these genes supported the finding that they are layer 4/5 markers during brain maturation.

Interestingly, *Inh L1-2 VIP TSPAN12* was previously reported to be found in layers 1-3 (Hodge et al., 2019; Langseth et al., 2021), however in this study the distribution of this cell type varies

with age. These VIP-expressing inhibitory neurons are believed to play a crucial role in regulating the balance of excitation and inhibition in sensory processing in the sensory cortex of mice (Mesik et al., 2015). Studies have demonstrated that VIP inhibitory interneurons are particularly susceptible to early neurodevelopmental disturbances, which may contribute to the development of neurodevelopmental disorders (Goff and Goldberg, 2021). The research findings suggest that alterations or malfunctioning of VIP inhibitory interneurons during development can impair the proper maturation of cortical circuits, leading to cognitive and behavioural deficits in adulthood (Goff and Goldberg, 2021). It is possible that the changes in the distribution of this cell type in this study could be attributed to the medical conditions of the samples, and therefore should be validated with further samples.

4.5 Validation of previously identified layer 1 neuronal marker genes, *CXCL14* and *RELN*, using Bayesspace and HCR

Using scRNA-seq Wu et al., identified *CXCL14* as an upper layer gene marker expressed by GABAergic inhibitory neurons, which is prominent in layer 1 of the human visual cortex (Wu et al., 2022). Using Bayesspace and HCR, this study confirmed that *CXCL14* is an upper layer marker in the temporal cortex and that it is highly enriched in layer 1 of both the paediatric and adult samples. Moreover, Bayesspace analysis suggests that the expression level of *CXCL14* may increase with brain maturation. In the adult mouse brain, *CXCL14* is expressed in many regions of the brain including the cortex and dentate gyrus (Banisadr et al., 2011). Banisadr et al., demonstrated that *CXCL14* was localized in dentate gyrus neurons that have GABAergic phenotype and expressed *GAD1* (Banisadr et al., 2011). This is consistent with Wu et al 2020's scRNA-seq study, where all the analysed cells expressed *GAD1*. This study confirmed the co-localization of *CXCL14* and layer 1 *GAD1* in the paediatric and adult samples. In another study, *CXCL14*-knockout mice showed behavioural abnormalities, such as a reduction of food intake during the postnatal stages (Tanegashima et al., 2010). Therefore, *CXCL14* could be involved in the establishment of neural circuits that are closely linked to feeding behaviour (Tanegashima et al., 2010).

RELN is a layer 1 marker gene that encodes for the Reelin protein. In the study conducted by Maynard et al., they were able to validate that *RELN* is a layer 1 marker using Visium spatial transcriptomics and smFISH in the adult human prefrontal cortex sample. In this study, Bayesspace and HCR was used to confirm this expression profile in both the paediatric and

adult temporal cortex samples. In contrast to the developing brain where Reelin is primarily produced by Cajal-Retzius cells, a specific type of GABAergic interneuron is the primary source of Reelin in the adult mouse brain, leading to a change in the distribution pattern of Reelin (Alcántara et al., 1998; Pesold et al., 1998). This change in Reelin distribution suggests that the functions of Reelin in the adult brain may differ from those in the developing brain. This could also be reason for the increase in expression of *RELN* in the adult brain suggested by the BayesSpcae results in this study. Reelin protein is involved in the proper migration of cortical neurons and regulates the various aspects of laminar formation in the developing human brain. Mutations in *RELN* are associated with several psychiatric disorders such as bipolar disorder, schizophrenia, and autistic spectrum disorder (Arioka et al., 2018). Furthermore, alterations in Reelin expression or signalling pathways may contribute to the development of epilepsy. Reelin expression was also found to be reduced in animal models of epilepsy (Gong et al., 2007), and mutations in genes involved in Reelin signalling have been associated with an increased risk of epilepsy in humans (Jossin, 2020). The lower *RELN* expression in the paediatric samples observed in this study may be attributed to the fact that the samples were obtained from patients with epilepsy. Further validation of these findings is needed through future studies that quantify the changes in the *RELN* expression profiles using additional paediatric and adult tissue samples when they are available.

Studying cell-cell communication is crucial for gaining a comprehensive understanding of physiological processes and their dysregulation in diseases, such as in epilepsy. Current methods, such as single-cell transcriptomics, have been used to investigate ligand-receptor interactions and cellular networks. For example, a recent study conducted by Kumar et al., (2022) focused on drug-resistant epilepsy and used single-cell cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) to reveal a pro-inflammatory microenvironment and direct physical interactions between microglia and T cells (Kumar et al., 2022). Another study by Buchin et al., (2022), which focused on temporal lobe epilepsy, used single-cell electrophysiology and morphology surveys of granule cells in human brain slices to examine how these properties change with disease progression. The study found that changes in granule cell excitability are linked to three ionic conductances (BK, Cav2.2, and Kir2.1) (Buchin et al., 2022). Gain of function in BK channels and alterations in Cav2.2 and Kir2.1 conductances were associated with increased network excitability. Using two

independent statistical methods, non-parametric Mann-Whitney U testing and random forests, they analysed the electrophysiology and morphology features of single cells and look at cellular properties associated with disease progression. These findings highlight the importance of considering gene expression profiles and potential gene regulatory networks in understanding epilepsy. Future experiments could use tools like HoloNet, which reconstructs cell-cell communication networks using spatial transcriptomics data, to provide further insights into tissue architecture and organization (Li et al., 2022). This approach has the potential to uncover previously unknown communication events and functional relationships between different cell types within tissues.

4.6 LIMITATIONS OF STUDY

There are limitations to this study that must be addressed. Firstly, the use of a low sample size, consisting of only four samples across three individuals restricted the study's ability to uncover robust patterns and draw definitive conclusions. Therefore, future experiments should add more samples to enhance the study's statistical power and generalizability.

Another limitation of this work is the lack of accounting for sex as a variable in the analysis. The neuroimaging study conducted by Bellis et al., (2001) revealed significant sex differences in brain maturation processes during childhood and adolescence (Bellis et al., 2001). Boys exhibited greater loss of gray matter (GM) volume and increases in both white matter (WM) and corpus callosum (CC) area compared to girls within a similar age range. These findings highlight the need to account for sex-related influences in experiments, as they can significantly impact brain development trajectories and contribute to a more comprehensive understanding of observed outcomes (Bellis et al., 2001). Future work should address this limitation by performing differential gene expression analyses that identifies genes that are uniquely expressed by each sex followed by pathway analyses.

Despite these limitations, this study showed that Visium is a powerful tool for exploring spatial gene expression profiles. Together with bioinformatics tools such as cell2location, Visium data can be used to reveal cell type assignments across a large tissue area. The different bioinformatic pipelines used in this study were able to provide spatial locations of cell types, abundances of cell types, gene expression signatures, as well as potential

expression trajectories during brain maturation. In the future the addition of more samples could be useful in verifying some of the trends seen.

4.7 CONCLUSION

This study provides valuable insights into the cellular composition and spatial distribution of cells in the human temporal cortex. The identification of subtypes of astrocytes and excitatory neurons and their laminar distributions, as well as the co-localization analysis, provides a more detailed understanding of the cellular diversity and complexity of the human brain. The differential expression analysis of genes in specific regions and cell types also highlights potential biomarkers and targets for further investigation into aging and neurodegenerative diseases. This study also highlighted the importance of confirming gene expression patterns using multiple methods (HCR and Bayesspace). However, the study is limited by its sample size and the lack of information on the functional roles of the identified cell types. Further studies are needed to validate and expand on these findings, as well as to investigate the functional implications of cellular heterogeneity in the human brain over the course of brain maturation.

REFERENCES

- Alcántara, S., Ruiz, M., D’Arcangelo, G., Ezan, F., De Lecea, L., Curran, T., Sotelo, C., Soriano, E., 1998. Regional and cellular patterns of reelin mRNA expression in the forebrain of the developing and adult mouse. *J. Neurosci.* 18, 7779–7799.
<https://doi.org/10.1523/jneurosci.18-19-07779.1998>
- AlJanahi, A.A., Danielsen, M., Dunbar, C.E., 2018. An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. *Mol. Ther. - Methods Clin. Dev.* 10, 189–196.
<https://doi.org/10.1016/j.omtm.2018.07.003>
- Arioka, Y., Shishido, E., Kubo, H., Kushima, I., Yoshimi, A., Kimura, H., Ishizuka, K., Aleksic, B., Maeda, T., Ishikawa, M., Kuzumaki, N., Okano, H., Mori, D., Ozaki, N., 2018. Single-cell trajectory analysis of human homogenous neurons carrying a rare RELN variant. *Transl. Psychiatry* 8. <https://doi.org/10.1038/s41398-018-0177-8>
- Asp, M., Bergenstråhle, J., Lundeberg, J., 2020. Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration. *BioEssays* 42, 1–16.
<https://doi.org/10.1002/bies.201900221>
- Bakken, T.E., van Velthoven, C.T., Menon, V., Hodge, R.D., Yao, Z., Nguyen, T.N., Graybuck, L.T., Horwitz, G.D., Bertagnolli, D., Goldy, J., Yanny, A.M., Garren, E., Parry, S., Casper, T., Ishehata, S., Barkan, E.R., Szafer, A., Levi, B.P., Dee, N., Smith, K.A., Sunkin, S.M., Bernard, A., Phillips, J., Hawrylycz, M.J., Koch, C., Murphy, G.J., Lein, E., Zeng, H., Tasic, B., 2021. Single-cell and single-nucleus RNA-seq uncovers shared and distinct axes of variation in dorsal LGN neurons in mice, non-human primates, and humans. *Elife* 10, 1–20. <https://doi.org/10.7554/eLife.64875>
- Banisadr, G., Frederick, T.J., Freitag, C., Ren, D., Jung, H., Miller, S.D., Miller, R.J., 2011. The role of CXCR4 signaling in the migration of transplanted oligodendrocyte progenitors into the cerebral white matter. *Neurobiol. Dis.* 44, 19–27.
<https://doi.org/10.1016/j.nbd.2011.05.019>
- Bellis, M.D. De, Keshavan, M.S., Beers, S.R., Hall, J., Frustaci, K., Masalehdan, A., Noll, J., Boring, A.M., 2001. Sex Differences in Brain Maturation during Childhood and Adolescence 552–557.

- Berglund, E., Saarenpää, S., Jemt, A., Gruselius, J., Larsson, L., Bergenstråhle, L., Lundeberg, J., Giacomello, S., 2020. Automation of Spatial Transcriptomics library preparation to enable rapid and robust insights into spatial organization of tissues. *BMC Genomics* 21, 1–10. <https://doi.org/10.1186/s12864-020-6631-z>
- Bradl, M., Lassmann, H., 2010. Oligodendrocytes: Biology and pathology. *Acta Neuropathol.* 119, 37–53. <https://doi.org/10.1007/s00401-009-0601-5>
- Buchin, A., Nandi, A., Koch, C., Ting, J., Anastassiou, C.A., Frates, R. De, Nandi, A., Mann, R., Chong, P., Ng, L., Miller, J., 2022. Article Multi-modal characterization and simulation of human epileptic circuitry II II Multi-modal characterization and simulation of human epileptic circuitry. *CellReports* 41, 111873. <https://doi.org/10.1016/j.celrep.2022.111873>
- Budday, S., Steinmann, P., Kuhl, E., 2015. Physical biology of human brain development. *Front. Cell. Neurosci.* 9, 1–17. <https://doi.org/10.3389/fncel.2015.00257>
- Bugiani, M., Plug, B.C., Man, J.H.K., Breur, M., van der Knaap, M.S., 2022. Heterogeneity of white matter astrocytes in the human brain. *Acta Neuropathol.* 143, 159–177. <https://doi.org/10.1007/s00401-021-02391-3>
- Cao, Y., Wang, X., Peng, G., 2019. SCSA: a cell type annotation tool for single-cell RNA-seq data. *bioRxiv* 2019.12.22.886481. <https://doi.org/10.1101/2019.12.22.886481>
- Cardozo, P.L., de Lima, I.B.Q., Maciel, E.M.A., Silva, N.C., Dobransky, T., Ribeiro, F.M., 2019. Synaptic Elimination in Neurological Disorders. *Curr. Neuropharmacol.* 17, 1071–1095. <https://doi.org/10.2174/1570159x17666190603170511>
- Carter, R.A., Bihannic, L., Rosencrance, C., Easton, J., Northcott, P.A., Gawad, C., Carter, R.A., Bihannic, L., Rosencrance, C., Hadley, J.L., Tong, Y., Phoenix, T.N., Natarajan, S., Easton, J., Northcott, P.A., Gawad, C., 2018. Article A Single-Cell Transcriptional Atlas of the Developing Article A Single-Cell Transcriptional Atlas of the Developing Murine Cerebellum. *Curr. Biol.* 28, 2910–2920.e2. <https://doi.org/10.1016/j.cub.2018.07.062>
- Chen, G., Ning, B., Shi, T., 2019. Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* 10, 1–13. <https://doi.org/10.3389/fgene.2019.00317>

- Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., Zhuang, X., 2015. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* (80-.). 348, 1360–1363. <https://doi.org/10.1126/science.aaa6090>
- Chen, W.T., Lu, A., Craessaerts, K., Pavie, B., Sala Frigerio, C., Corthout, N., Qian, X., Laláková, J., Kühnemund, M., Voytyuk, I., Wolfs, L., Mancuso, R., Salta, E., Balusu, S., Snellinx, A., Munck, S., Jurek, A., Fernandez Navarro, J., Saido, T.C., Huitinga, I., Lundeborg, J., Fiers, M., De Strooper, B., 2020. Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer’s Disease. *Cell* 182, 976–991.e19. <https://doi.org/10.1016/j.cell.2020.06.038>
- Choi, H.M.T., Calvert, C.R., Husain, N., Huss, D., Barsi, J.C., Deverman, B.E., Hunter, R.C., Kato, M., Lee, S.M., Abelin, A.C.T., Rosenthal, A.Z., Akbari, O.S., Li, Y., Hay, B.A., Sternberg, P.W., Patterson, P.H., Davidson, E.H., Mazmanian, S.K., Prober, D.A., Van De Rijn, M., Leadbetter, J.R., Newman, D.K., Readhead, C., Bronner, M.E., Wold, B., Lansford, R., Sauka-Spengler, T., Fraser, S.E., Pierce, N.A., 2016. Mapping a multiplexed zoo of mRNA expression. *Dev.* 143, 3632–3637. <https://doi.org/10.1242/dev.140137>
- Clarke, L.E., Liddelow, S.A., Chakraborty, C., Münch, A.E., Heiman, M., Barres, B.A., 2018. Normal aging induces A1-like astrocyte reactivity. *Proc. Natl. Acad. Sci. U. S. A.* 115, E1896–E1905. <https://doi.org/10.1073/pnas.1800165115>
- Codeluppi, S., Borm, L.E., Zeisel, A., La Manno, G., van Lunteren, J.A., Svensson, C.I., Linnarsson, S., 2018. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* 15, 932–935. <https://doi.org/10.1038/s41592-018-0175-z>
- Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Gephart, M.G.H., Barres, B.A., Quake, S.R., 2015. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7285–7290. <https://doi.org/10.1073/pnas.1507125112>
- Dorrity, M.W., Saunders, L.M., Queitsch, C., Fields, S., Trapnell, C., 2020. Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nat. Commun.* 11, 1–6. <https://doi.org/10.1038/s41467-020-15351-4>
- Ecker, J.R., Geschwind, D.H., Kriegstein, A.R., Ngai, J., Osten, P., Polioudakis, D., Regev, A., Sestan, N., Wickersham, I.R., Zeng, H., 2017a. The BRAIN Initiative Cell Census

- Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. *Neuron* 96, 542–557. <https://doi.org/10.1016/j.neuron.2017.10.007>
- Ecker, J.R., Geschwind, D.H., Kriegstein, A.R., Ngai, J., Osten, P., Polioudakis, D., Regev, A., Sestan, N., Wickersham, I.R., Zeng, H., 2017b. The BRAIN Initiative Cell Census Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. *Neuron* 96, 542–557. <https://doi.org/10.1016/j.neuron.2017.10.007>
- Eng, C.H.L., Lawson, M., Zhu, Q., Dries, R., Koulana, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.C., Cai, L., 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 568, 235–239. <https://doi.org/10.1038/s41586-019-1049-y>
- Evans, S.J., Choudary, P. V., Vawter, M.P., Li, J., Meador-Woodruff, J.H., Lopez, J.F., Burke, S.M., Thompson, R.C., Myers, R.M., Jones, E.G., Bunney, W.E., Watson, S.J., Akil, H., 2003. DNA microarray analysis of functionally discrete human brain regions reveals divergent transcriptional profiles. *Neurobiol. Dis.* 14, 240–250. [https://doi.org/10.1016/S0969-9961\(03\)00126-8](https://doi.org/10.1016/S0969-9961(03)00126-8)
- Evans, S.J., Datson, N.A., Kabbaj, M., Thompson, R.C., Vreugdenhil, E., De Kloet, E.R., Watson, S.J., Akil, H., 2002. Evaluation of Affymetrix Gene Chip sensitivity in rat hippocampal tissue using SAGE analysis. *Eur. J. Neurosci.* 16, 409–413. <https://doi.org/10.1046/j.1460-9568.2002.02097.x>
- Eze, U.C., Bhaduri, A., Haeussler, M., Nowakowski, T.J., Kriegstein, A.R., 2021. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat. Neurosci.* 24, 584–594. <https://doi.org/10.1038/s41593-020-00794-1>
- Fan, X., Dong, J., Zhong, S., Wei, Y., Wu, Q., Yan, L., Yong, J., Sun, L., Wang, Xiaoye, Zhao, Y., Wang, W., Yan, J., Wang, Xiaoqun, Qiao, J., Tang, F., 2018. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* 28, 730–745. <https://doi.org/10.1038/s41422-018-0053-3>
- Fang, R., Xia, C., Close, J.L., Zhang, M., He, J., Huang, Z., Halpern, A.R., Long, B., Miller, J.A., Lein, E.S., Zhuang, X., 2022. Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* (80-.). 377, 56–62.

<https://doi.org/10.1126/science.abm1741>

Farhy-Tselnicker, I., Boisvert, M.M., Liu, H., Dowling, C., Erikson, G.A., Blanco-Suarez, E., Farhy, C., Shokhirev, M.N., Ecker, J.R., Allen, N.J., 2021. Activity-dependent modulation of synapse-regulating genes in astrocytes. *Elife* 10, 1–43.

<https://doi.org/10.7554/eLife.70514>

Fawkner-Corbett, D., Antanaviciute, A., Parikh, K., Jagielowicz, M., Gerós, A.S., Gupta, T., Ashley, N., Khamis, D., Fowler, D., Morrissey, E., Cunningham, C., Johnson, P.R.V., Koohy, H., Simmons, A., 2021. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* 184, 810–826.e23.

<https://doi.org/10.1016/j.cell.2020.12.016>

Femino, A.M., Fay, F.S., Fogarty, K., Singer, R.H., Dwyer, P., 1998. Visualization of single RNA transcripts in situ. *Science* (80-.). 280, 585–590.

<https://doi.org/10.1126/science.280.5363.585>

Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., Solas, D., 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* (80-.). 251, 767–773.

<https://doi.org/10.1126/science.1990438>

Germann, M., Brederoo, S.G., Sommer, I.E.C., 2021. Abnormal synaptic pruning during adolescence underlying the development of psychotic disorders. *Curr. Opin. Psychiatry* 34, 222–227. <https://doi.org/10.1097/YCO.0000000000000696>

Goff, K.M., Goldberg, E.M., 2021. A Role for Vasoactive Intestinal Peptide Interneurons in Neurodevelopmental Disorders. *Dev. Neurosci.* 43, 168–180.

<https://doi.org/10.1159/000515264>

Gong, C., Wang, T.W., Huang, H.S., Parent, J.M., 2007. Reelin regulates neuronal progenitor migration in intact and epileptic hippocampus. *J. Neurosci.* 27, 1803–1811.

<https://doi.org/10.1523/JNEUROSCI.3111-06.2007>

Gonzalez, D.M., Gregory, J., Brennand, K.J., 2017. The importance of non-neuronal cell types in hiPSC-based disease modeling and drug screening. *Front. Cell Dev. Biol.* 5, 1–14.

<https://doi.org/10.3389/fcell.2017.00117>

- Grillner, S., Ip, N., Koch, C., Koroshetz, W., Okano, H., Polachek, M., Poo, M., Sejnowski, T.J., 2017. Worldwide initiatives to advance brain research. *Physiol. Behav.* 176, 139–148. <https://doi.org/10.1038/nn.4371.Worldwide>
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E.P., Jain, J., Srivastava, A., Stuart, T., Fleming, L.M., Yeung, B., Rogers, A.J., McElrath, J.M., Blish, C.A., Gottardo, R., Smibert, P., Satija, R., 2021. Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>
- Hasel, P., Rose, I.V.L., Sadick, J.S., Kim, R.D., Liddelow, S.A., 2021. Neuroinflammatory astrocyte subtypes in the mouse brain. *Nat. Neurosci.* 24, 1475–1487. <https://doi.org/10.1038/s41593-021-00905-6>
- Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., Yao, Z., Eggermont, J., Höllt, T., Levi, B.P., Shehata, S.I., Aevermann, B., Beller, A., Bertagnolli, D., Brouner, K., Casper, T., Cobbs, C., Dalley, R., Dee, N., Ding, S.L., Ellenbogen, R.G., Fong, O., Garren, E., Goldy, J., Gwinn, R.P., Hirschstein, D., Keene, C.D., Keshk, M., Ko, A.L., Lathia, K., Mahfouz, A., Maltzer, Z., McGraw, M., Nguyen, T.N., Nyhus, J., Ojemann, J.G., Oldre, A., Parry, S., Reynolds, S., Rimorin, C., Shapovalova, N. V., Somasundaram, S., Szafer, A., Thomsen, E.R., Tieu, M., Quon, G., Scheuermann, R.H., Yuste, R., Sunkin, S.M., Lelieveldt, B., Feng, D., Ng, L., Bernard, A., Hawrylycz, M., Phillips, J.W., Tasic, B., Zeng, H., Jones, A.R., Koch, C., Lein, E.S., 2019. Conserved cell types with divergent features in human versus mouse cortex. *Nature* 573, 61–68. <https://doi.org/10.1038/s41586-019-1506-7>
- Hou, X., Yang, Y., Li, P., Zeng, Z., Hu, W., Zhe, R., Liu, X., Tang, D., Ou, M., Dai, Y., 2021. Integrating Spatial Transcriptomics and Single-Cell RNA-seq Reveals the Gene Expression Profiling of the Human Embryonic Liver. *Front. Cell Dev. Biol.* 9, 1–14. <https://doi.org/10.3389/fcell.2021.652408>
- Hu, J., Li, X., Coleman, K., Schroeder, A., Irwin, D.J., Lee, E.B., Li, M., 2020. Integrating gene expression, spatial location and histology to identify spatial 1 domains and spatially variable genes by graph convolutional network. *bioRxiv* 2020.11.30.405118.

- Hu, J., Schroeder, A., Coleman, K., Chen, C., Auerbach, B.J., Li, M., 2021. Statistical and machine learning methods for spatially resolved transcriptomics with histology. *Comput. Struct. Biotechnol. J.* 19, 3829–3841.
<https://doi.org/10.1016/j.csbj.2021.06.052>
- Hwang, B., Lee, J.H., Bang, D., 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 50. <https://doi.org/10.1038/s12276-018-0071-8>
- Jacobsen, L.K., Picciotto, M.R., Heath, C.J., Mencl, W.E., Gelernter, J., 2009. Allelic Variation of Calsyntenin 2 (CLSTN2) Modulates the Impact of Developmental Tobacco Smoke Exposure on Mnemonic Processing in Adolescents. *Biol. Psychiatry* 65, 671–679.
<https://doi.org/10.1016/j.biopsych.2008.10.024>
- Joglekar, A., Prijbelski, A., Mahfouz, A., Collier, P., Lin, S., Schlusche, A.K., Marrocco, J., Williams, S.R., Haase, B., Hayes, A., Chew, J.G., Weisenfeld, N.I., Wong, M.Y., Stein, A.N., Hardwick, S.A., Hunt, T., Wang, Q., Dieterich, C., Bent, Z., Fedrigo, O., Sloan, S.A., Risso, D., Jarvis, E.D., Flicek, P., Luo, W., Pitt, G.S., Frankish, A., Smit, A.B., Ross, M.E., Tilgner, H.U., 2021. A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat. Commun.* 12, 1–16.
<https://doi.org/10.1038/s41467-020-20343-5>
- Jolly, S., Lang, V., Koelzer, V.H., Sala Frigerio, C., Magno, L., Salinas, P.C., Whiting, P., Palomer, E., 2019. Single-Cell Quantification of mRNA Expression in The Human Brain. *Sci. Rep.* 9, 1–9. <https://doi.org/10.1038/s41598-019-48787-w>
- Jossin, Y., 2020. Reelin functions, mechanisms of action and signaling pathways during brain development and maturation. *Biomolecules* 10, 1–31.
<https://doi.org/10.3390/BIOM10060964>
- Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M.M., Pletikos, M., Meyer, K.A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M.B., Krsnik, Ž., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S.N., Vortmeyer, A., Weinberger, D.R., Mane, S., Hyde, T.M., Huttner, A., Reimers, M., Kleinman, J.E., Šestan, N., 2011. Spatio-temporal transcriptome of the human brain. *Nature* 478, 483–489.

<https://doi.org/10.1038/nature10523>

- Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., von Feilitzen, K., Zwahlen, M., Arif, M., Altay, O., Li, X., Ozcan, M., Mardonoglu, A., Fagerberg, L., Mulder, J., Luo, Y., Ponten, F., Uhlén, M., Lindskog, C., 2021. A single-cell type transcriptomics map of human tissues. *Sci. Adv.* 7, 1–10. <https://doi.org/10.1126/sciadv.abh2169>
- Kim, M., Radaelli, C., Thomsen, E.R., Machen, D., Chartrand, T., Jorstad, N.L., 2022. Target cell-specific synaptic dynamics of excitatory to inhibitory neuron connections in supragranular layers of human neocortex.
- Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., Gayoso, A., Jain, M.S., Park, J.S., Ramona, L., Tuck, E., Arutyunyan, A., Vento-Tormo, R., Gerstung, M., James, L., Stegle, O., Bayraktar, O.A., 2022. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-01139-4>
- Knuesel, I., Chicha, L., Britschgi, M., Schobel, S.A., Bodmer, M., Hellings, J.A., Toovey, S., Prinssen, E.P., 2014. Maternal immune activation and abnormal brain development across CNS disorders. *Nat. Rev. Neurol.* 10, 643–660. <https://doi.org/10.1038/nrneurol.2014.187>
- Kokkat, T.J., Patel, M.S., McGarvey, D., Livolsi, V.A., Baloch, Z.W., 2013. Archived formalin-fixed paraffin-embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA, and protein. *Biopreserv. Biobank.* 11, 101–106. <https://doi.org/10.1089/bio.2012.0052>
- Kukurba, K.R., Montgomery, S.B., 2015. RNA sequencing and analysis. *Cold Spring Harb. Protoc.* 2015, 951–969. <https://doi.org/10.1101/pdb.top084970>
- Kumar, P., Lim, A., Hazirah, S.N., Jian, C., Chua, H., Ngoh, A., Poh, S.L., Yeo, T.H., Lim, J., Ling, S., Sutamam, N.B., Petretto, E., Chyi, D., Low, Y., Zeng, L., Tan, E., Arkachaisri, T., Yeo, J.G., Ginhoux, F., Chan, D., Albani, S., 2022. Single-cell transcriptomics and surface epitope detection in human brain epileptic lesions identifies pro-inflammatory signaling 25. <https://doi.org/10.1038/s41593-022-01095-5>

- Kumar, V., Krolewski, D.M., Hebda-Bauer, E.K., Parsegian, A., Martin, B., Foltz, M., Akil, H., Watson, S.J., 2021. Optimization and evaluation of fluorescence in situ hybridization chain reaction in cleared fresh-frozen brain tissues. *Brain Struct. Funct.* 226, 481–499. <https://doi.org/10.1007/s00429-020-02194-4>
- Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.L., Chen, S., Vijayaraghavan, R., Wong, J., Chen, A., Sheng, X., Kaper, F., Shen, R., Ronaghi, M., Fan, J.B., Wang, W., Chun, J., Zhang, K., 2016. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* (80-.). 352, 1586–1590. <https://doi.org/10.1126/science.aaf1204>
- Langseth, C.M., Gyllborg, D., Miller, J.A., Close, J.L., Long, B., Lein, E.S., Hilscher, M.M., Nilsson, M., 2021. Comprehensive in situ mapping of human cortical transcriptomic cell types. *Commun. Biol.* 4, 1–7. <https://doi.org/10.1038/s42003-021-02517-z>
- Laukka, E.J., Lövdén, M., Herlitz, A., Karlsson, S., Ferencz, B., Pantzar, A., Keller, L., Graff, C., Fratiglioni, L., Bäckman, L., 2013. Genetic effects on old-age cognitive functioning: A population-based study. *Psychol. Aging* 28, 262–274. <https://doi.org/10.1037/a0030829>
- Lein, E., Borm, L.E., Linnarsson, S., 2017. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* (80-.). 358, 64–69. <https://doi.org/10.1126/science.aan6827>
- Lepeta, K., Lourenco, M. V., Schweitzer, B.C., Martino Adami, P. V., Banerjee, P., Catuara-Solarz, S., de La Fuente Revenga, M., Guillem, A.M., Haidar, M., Ijomone, O.M., Nadorp, B., Qi, L., Perera, N.D., Refsgaard, L.K., Reid, K.M., Sabbar, M., Sahoo, A., Schaefer, N., Sheean, R.K., Suska, A., Verma, R., Vicidomini, C., Wright, D., Zhang, X.D., Seidenbecher, C., 2016. Synaptopathies: synaptic dysfunction in neurological disorders – A review from students to students. *J. Neurochem.* 785–805. <https://doi.org/10.1111/jnc.13713>
- Li, H., Ma, T., Hao, M., Wei, L., Zhang, X., 2022. Decoding functional cell-cell communication events by multi-view graph learning on spatial transcriptomics. *bioRxiv* 2022.06.22.496105.
- Li, P.Y., Wang, P., Gao, S.G., Dong, D.Y., 2020. Long noncoding RNA SOX2-OT: Regulations,

- functions, and roles on mental illnesses, cancers, and diabetic complications. *Biomed Res. Int.* 2020. <https://doi.org/10.1155/2020/2901589>
- Li, X., Wang, C.Y., 2021. From bulk, single-cell to spatial RNA sequencing. *Int. J. Oral Sci.* 13, 1–6. <https://doi.org/10.1038/s41368-021-00146-0>
- Liao, J., Lu, X., Shao, X., Zhu, L., Fan, X., 2021. Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends Biotechnol.* 39, 43–58. <https://doi.org/10.1016/j.tibtech.2020.05.006>
- Lipina, T. V., Prasad, T., Yokomaku, D., Luo, L., Connor, S.A., Kawabe, H., Wang, Y.T., Brose, N., Roder, J.C., Craig, A.M., 2016. Cognitive Deficits in Calsyntenin-2-deficient Mice Associated with Reduced GABAergic Transmission. *Neuropsychopharmacology* 41, 802–810. <https://doi.org/10.1038/npp.2015.206>
- Liu, L., Feigin, V., Sacco, R.L., Koroshetz, W.J., 2020. Promoting global collaboration for brain health research. *BMJ* 371, 1–4. <https://doi.org/10.1136/bmj.m3753>
- Liu, Y., Yang, M., Deng, Y., Su, G., Enniful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., Norris, E., Pan, A., Li, J., Xiao, Y., Halene, S., Fan, R., 2020. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* 183, 1665–1681.e18. <https://doi.org/10.1016/j.cell.2020.10.026>
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T., 2017. Transcriptomics technologies. *PLoS Comput. Biol.* 13, 1–23. <https://doi.org/10.1371/journal.pcbi.1005457>
- Mantri, M., Scuderi, G.J., Abedini-Nassab, R., Wang, M.F.Z., McKellar, D., Shi, H., Grodner, B., Butcher, J.T., De Vlaminc, I., 2021. Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis. *Nat. Commun.* 12. <https://doi.org/10.1038/s41467-021-21892-z>
- Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippani, M., Chew, J., Yin, Y., Kleinman, J.E., Hyde, T.M., Rao, N., Hicks, S.C., Martinowich, K., Jaffe, A.E., 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* 24. <https://doi.org/10.1038/s41593-020-00787-0>

- Mesik, L., Ma, W.P., Li, L.Y., Ibrahim, L.A., Huang, Z.J., Zhang, L., Tao, H.W., 2015. Functional response properties of VIP-expressing inhibitory neurons in mouse visual and auditory cortex. *Front. Neural Circuits* 9, 1–14. <https://doi.org/10.3389/fncir.2015.00022>
- Mohan, H., Verhoog, M.B., Doreswamy, K.K., Eyal, G., Aardse, R., Lodder, B.N., Goriounova, N.A., Asamoah, B., Brakspear, A.B.C., Groot, C., Van Der Sluis, S., Testa-Silva, G., Obermayer, J., Boudewijns, Z.S.R.M., Narayanan, R.T., Baayen, J.C., Segev, I., Mansvelder, H.D., De Kock, C.P.J., 2015. Dendritic and axonal architecture of individual pyramidal neurons across layers of adult human neocortex. *Cereb. Cortex* 25, 4839–4853. <https://doi.org/10.1093/cercor/bhv188>
- Molliver, M.E., Kostović, I., Van Der Loos, H., 1973. The development of synapses in cerebral cortex of the human fetus. *Brain Res.* 50, 403–407. [https://doi.org/10.1016/0006-8993\(73\)90741-5](https://doi.org/10.1016/0006-8993(73)90741-5)
- Moreno-García, A., Kun, A., Calero, O., Medina, M., Calero, M., 2018. An overview of the role of lipofuscin in age-related neurodegeneration. *Front. Neurosci.* 12, 1–13. <https://doi.org/10.3389/fnins.2018.00464>
- Needham, H., Torpey, G., Flores, C.C., Davis, C.J., Vanderheyden, W.M., Gerstner, J.R., 2022. A Dichotomous Role for FABP7 in Sleep and Alzheimer’s Disease Pathogenesis: A Hypothesis. *Front. Neurosci.* 16, 1–19. <https://doi.org/10.3389/fnins.2022.798994>
- Orsenigo, F., Conze, L.L., Jauhainen, S., Corada, M., Lazzaroni, F., Malinverno, M., Sundell, V., Cunha, S.I., Brännström, J., Globisch, M.A., Maderna, C., Lampugnani, M.G., Magnusson, P.U., Dejana, E., 2020. Mapping endothelial-cell diversity in cerebral cavernous malformations at single-cell resolution. *Elife* 9, 1–34. <https://doi.org/10.7554/eLife.61413>
- Pan, J., Ma, N., Yu, B., Zhang, W., Wan, J., 2020. Transcriptomic profiling of microglia and astrocytes throughout aging. *J. Neuroinflammation* 17, 1–19. <https://doi.org/10.1186/s12974-020-01774-9>
- Pesold, C., Impagnatiello, F., Pisu, M.G., Uzunov, D.P., Costa, E., Guidotti, A., Caruncho, H.J., 1998. Reelin is preferentially expressed in neurons synthesizing γ -aminobutyric acid in cortex and hippocampus of adult rats. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3221–3226.

<https://doi.org/10.1073/pnas.95.6.3221>

- Pfisterer, U., Petukhov, V., Demharter, S., Meichsner, J., Thompson, J.J., Batiuk, M.Y., Martinez, A.A., Vasistha, N.A., Thakur, A., Mikkelsen, J., Adorjan, I., Pinborg, L.H., Pers, T.H., von Engelhardt, J., Kharchenko, P. V., Khodosevich, K., 2020. Identification of epilepsy-associated neuronal subtypes and gene expression underlying epileptogenesis. *Nat. Commun.* 11. <https://doi.org/10.1038/s41467-020-18752-7>
- Qi, C., Luo, L. Da, Feng, I., Ma, S., 2022. Molecular mechanisms of synaptogenesis. *Front. Synaptic Neurosci.* 14, 1–20. <https://doi.org/10.3389/fnsyn.2022.939793>
- Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., Tyagi, S., 2008. Imaging individual mRNA molecules using single labeled probes. *Nat. Methods* 5, 1–40. <https://doi.org/10.1038/nmeth.1253>
- Regev, A., Teichmann, S., Lander, E., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Enard, W., 2017. Science Forum: The Human Cell Atlas. *Elife* 1–30.
- Ren, X., Zhong, G., Zhang, Q., Zhang, L., Sun, Y., Zhang, Z., 2020. Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res.* 30, 763–778. <https://doi.org/10.1038/s41422-020-0353-2>
- Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., Macosko, E.Z., 2019. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* (80-.). 363, 1463–1467. <https://doi.org/10.1126/science.aaw1219>
- Roth, R., Kim, S., Kim, J., Rhee, S., 2020. Single-cell and spatial transcriptomics approaches of cardiovascular development and disease. *BMB Rep.* 53, 393–399. <https://doi.org/10.5483/BMBRep.2020.53.8.130>
- Rudolf, G., Lesca, G., Mehrjouy, M.M., Labalme, A., Salmi, M., Bache, I., Bruneau, N., Pendziwiat, M., Fluss, J., De Bellescize, J., Scholly, J., Møller, R.S., Craiu, D., Tommerup, N., Valenti-Hirsch, M.P., Schluth-Bolard, C., Sloan-Béna, F., Helbig, K.L., Weckhuysen, S., Edery, P., Coulbaut, S., Abbas, M., Scheffer, I.E., Tang, S., Myers, C.T., Stamberger, H.,

- Carvill, G.L., Shinde, D.N., Mefford, H.C., Neagu, E., Huether, R., Lu, H.M., Dica, A., Cohen, J.S., Iliescu, C., Pomeran, C., Rubenstein, J., Helbig, I., Sanlaville, D., Hirsch, E., Szepetowski, P., 2016. Loss of function of the retinoid-related nuclear receptor (RORB) gene and epilepsy. *Eur. J. Hum. Genet.* 24, 1761–1770.
<https://doi.org/10.1038/ejhg.2016.80>
- Sampaio-Baptista, C., Vallès, A., Khrapitchev, A.A., Akkermans, G., Winkler, A.M., Foxley, S., Sibson, N.R., Roberts, M., Miller, K., Diamond, M.E., Martens, G.J.M., De Weerd, P., Johansen-Berg, H., 2020. White matter structure and myelin-related gene expression alterations with experience in adult rats. *Prog. Neurobiol.* 187, 101770.
<https://doi.org/10.1016/j.pneurobio.2020.101770>
- Scaduto, P., Lauterborn, J.C., Cox, C.D., Fracassi, A., Zeppillo, T., Gutierrez, B.A., Keene, C.D., Crane, P.K., Mukherjee, S., Russell, W.K., Taglialatela, G., Limon, A., 2022. Functional excitatory to inhibitory synaptic imbalance and loss of cognitive performance in people with Alzheimer’s disease neuropathologic change. *Acta Neuropathol.* 145, 303–324.
<https://doi.org/10.1007/s00401-022-02526-0>
- Schena, M., Shalon, D., Davis, R. and Brown, P., 1995. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray Author (s): Mark Schena , Dari Shalon , Ronald W . Davis , Patrick O . Brown Published by : American Association for the Advancement of Science Stable URL : <http://w. Adv. Sci. 270, 467–470>.
- Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., Ragg, T., 2006. The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* 7, 1–14.
<https://doi.org/10.1186/1471-2199-7-3>
- Sedmak, G., Judaš, M., 2021. White matter interstitial neurons in the adult human brain: 3% of cortical neurons in quest for recognition. *Cells* 10, 1–13.
<https://doi.org/10.3390/cells10010190>
- Sekar, A., Bialas, A.R., De Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., Genovese, G., Rose, S.A., Handsaker, R.E., Daly, M.J., Carroll, M.C., Stevens, B., McCarroll, S.A., 2016. Schizophrenia risk from complex

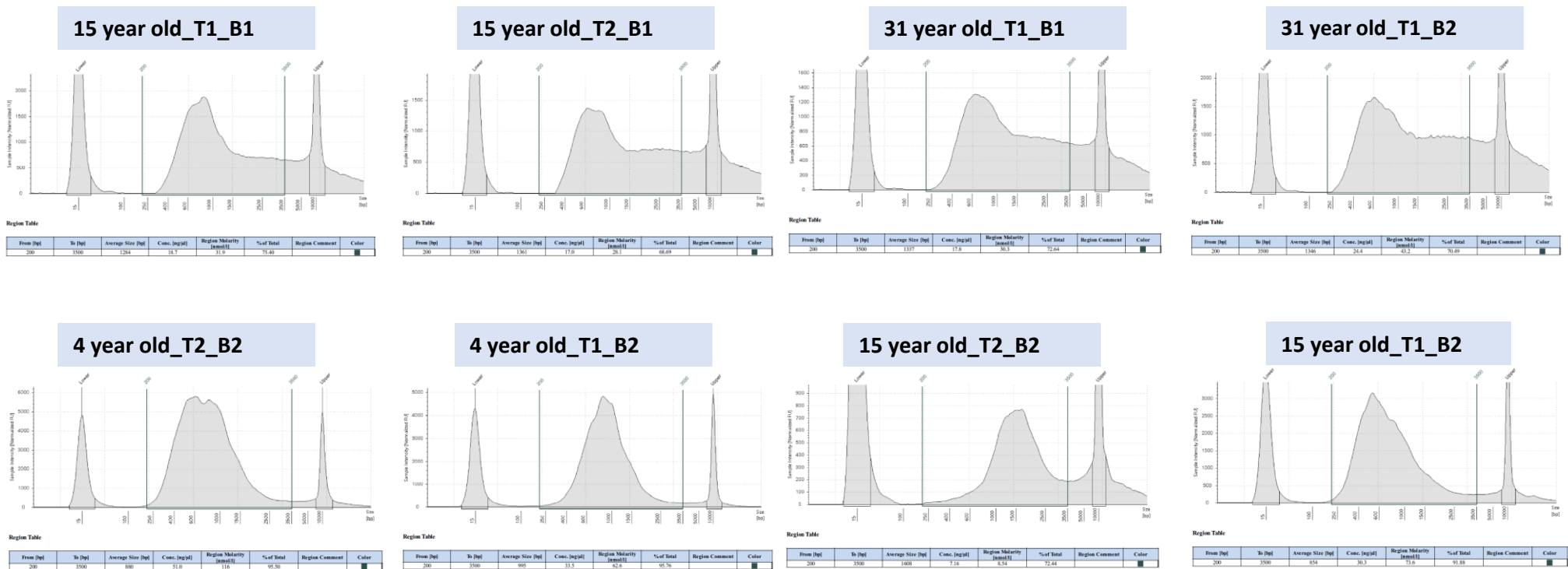
- variation of complement component 4. *Nature* 530, 177–183.
<https://doi.org/10.1038/nature16549>
- Shah, S., Lubeck, E., Schwarzkopf, M., He, T.F., Greenbaum, A., Sohn, C.H., Lignell, A., Choi, H.M.T., Gradinaru, V., Pierce, N.A., Cai, L., 2016a. Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Dev.* 143, 2862–2867. <https://doi.org/10.1242/dev.138560>
- Shah, S., Lubeck, E., Zhou, W., Cai, L., 2016b. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* 92, 342–357. <https://doi.org/10.1016/j.neuron.2016.10.001>
- Silbereis, J.C., Pochareddy, S., Zhu, Y., Li, M., Sestan, N., 2017. 乳鼠心肌提取 HHS Public Access. *Physiol. Behav.* 176, 139–148.
<https://doi.org/10.1016/j.neuron.2015.12.008>
- Somiari, S.B., Shuss, S., Liu, J., Mamula, K., O'Donnell, A., Deyarmin, B., Kane, J., Greenawalt, A., Larson, C., Rigby, S., Hu, H., Shriver, C.D., 2022. Assessing the quality of RNA isolated from human breast tissue after ambient room temperature exposure. *PLoS One* 17, 1–12. <https://doi.org/10.1371/journal.pone.0262654>
- Song, L., Pan, S., Zhang, Z., Jia, L., Chen, W.H., Zhao, X.M., 2021. STAB: A spatio-temporal cell atlas of the human brain. *Nucleic Acids Res.* 49, D1029–D1037.
<https://doi.org/10.1093/nar/gkaa762>
- Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P.I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J., Frisén, J., 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* (80-.). 353, 78–82. <https://doi.org/10.1126/science.aaf2403>
- Stiles, J., Jernigan, T.L., 2010. The basics of brain development. *Neuropsychol. Rev.* 20, 327–348. <https://doi.org/10.1007/s11065-010-9148-4>
- Stortchevoi, A., Kamelamela, N., Levine, S.S., 2020. SPRI beads-based size selection in the range of 2-10kb. *J. Biomol. Tech.* 31, 7–10. <https://doi.org/10.7171/jbt.20-3101-002>

- Sun, Y., Zhang, L., Johnston, N.L., Torrey, E.F., Yolken, R.H., 2001. Serial analysis of gene expression in the frontal cortex of patients with bipolar disorder. *Br. J. Psychiatry* 178, 13–17. <https://doi.org/10.1192/bjp.178.41.s137>
- Tanegashima, K., Okamoto, S., Nakayama, Y., Taya, C., Shitara, H., Ishii, R., Yonekawa, H., Minokoshi, Y., Hara, T., 2010. CXCL14 deficiency in mice attenuates obesity and inhibits feeding behavior in a novel environment. *PLoS One* 5, 1–9. <https://doi.org/10.1371/journal.pone.0010321>
- 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, 377–382. <https://doi.org/10.1038/nmeth.1315>
- Tang, X., Huang, Y., Lei, J., Luo, H., Zhu, X., 2019. The single-cell sequencing: New developments and medical applications. *Cell Biosci.* 9, 1–9. <https://doi.org/10.1186/s13578-019-0314-y>
- Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K.E., Lathia, K., Rimorin, C., Tieu, M., Larsen, R., Casper, T., Barkan, E., Kroll, M., Parry, S., Shapovalova, N. V., Hirschstein, D., Pendergraft, J., Sullivan, H.A., Kim, T.K., Szafer, A., Dee, N., Groblewski, P., Wickersham, I., Cetin, A., Harris, J.A., Levi, B.P., Sunkin, S.M., Madisen, L., Daigle, T.L., Looger, L., Bernard, A., Phillips, J., Lein, E., Hawrylycz, M., Svoboda, K., Jones, A.R., Koch, C., Zeng, H., 2018. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563, 72–78. <https://doi.org/10.1038/s41586-018-0654-5>
- Tau, G.Z., Peterson, B.S., 2010. Normal development of brain circuits. *Neuropsychopharmacology* 35, 147–168. <https://doi.org/10.1038/npp.2009.115>
- Teves, J.M., Won, K.J., 2020. Mapping cellular coordinates through advances in spatial transcriptomics technology. *Mol. Cells* 43, 591–599. <https://doi.org/10.14348/molcells.2020.0020>
- Tippani, M., Divecha, H.R., Catallini, J.L., Weber, L.M., Spangler, A., Jaffe, A.E., Hicks, S.C., Martinowich, K., Collado-Torres, L., Page, S.C., Maynard, K.R., 2021. VistoSeg: a

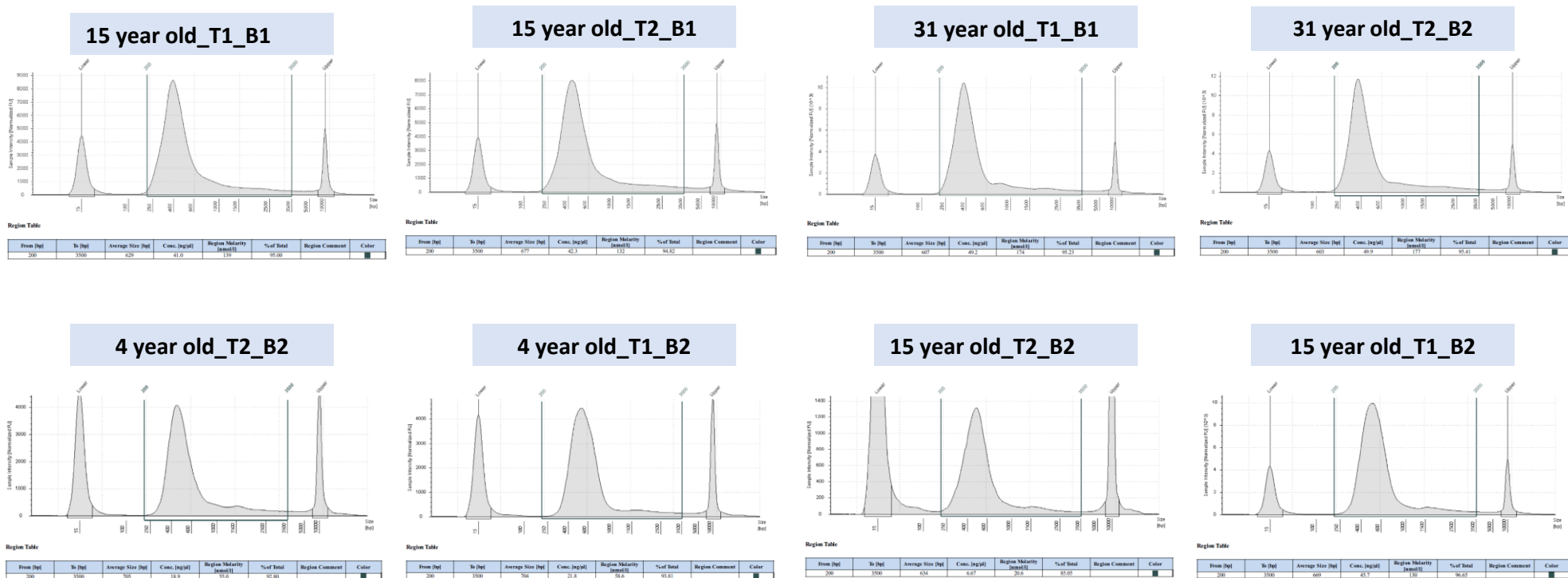
- MATLAB pipeline to process, analyze and visualize high-resolution histology images for Visium spatial transcriptomics data. bioRxiv. <https://doi.org/10.5281/zenodo.5156783>
- Tosetti, V., Sassone, J., Ferri, A.L.M., Taiana, M., Bedini, G., Nava, S., Brenna, G., Di Resta, C., Pareyson, D., Giulio, A.M. Di, Carelli, S., Parati, E.A., Gorio, A., 2017. Transcriptional role of androgen receptor in the expression of long non-coding RNA Sox2OT in neurogenesis. *PLoS One* 12, 1–20. <https://doi.org/10.1371/journal.pone.0180579>
- Trapnell, C., Liu, S., 2016. Single-cell transcriptome sequencing: Recent advances and remaining challenges. *F1000Research* 5. <https://doi.org/10.12688/f1000research.7223.1>
- V ilchez-Acosta, A., Manso, Y., Cardenas, A., Elias-Tersa, A., Martinez-Losa, M., Pascual, M., Alvarez-Dolado, M., Nairn, A.C., Borrell, V., Soriano, E., 2022. Specific contribution of Reelin expressed by Cajal – Retzius cells or GABAergic interneurons to cortical lamination 1–12. <https://doi.org/10.1073/pnas.2120079119/-/DCSupplemental>. Published
- Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial Analysis of Gene Expression. *Science* 270, 484–487. doi:10.1126/science.270.5235.484
- Verkhatsky, A., Augusto-Oliveira, M., Pivoriūnas, A., Popov, A., Brazhe, A., Semyanov, A., 2021. Astroglial asthenia and loss of function, rather than reactivity, contribute to the ageing of the brain. *Pflugers Arch. Eur. J. Physiol.* 473, 753–774. <https://doi.org/10.1007/s00424-020-02465-3>
- Wang, G., Ang, C.-E., Fan, J., Wang, A., Moffitt, J.R., Zhuang, X., 2020. Spatial organization of the transcriptome in individual neurons. bioRxiv 2020.12.07.414060.
- Wang, G., Moffitt, J.R., Zhuang, X., 2018. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep.* 8, 1–13. <https://doi.org/10.1038/s41598-018-22297-7>
- Wang, J., Devlin, B., Roeder, K., 2020. Using multiple measurements of tissue to estimate subject- and cell-type-specific gene expression. *Bioinformatics* 36, 782–788. <https://doi.org/10.1093/bioinformatics/btz619>

- Wang, W., Wang, G.Z., 2019. Understanding molecular mechanisms of the brain through transcriptomics. *Front. Physiol.* 10, 1–8. <https://doi.org/10.3389/fphys.2019.00214>
- Wang, X., He, Y., Zhang, Q., Ren, X., Zhang, Z., 2019. Direct Comparative Analysis of 10X Genomics Chromium and Smart-seq2. *bioRxiv* 615013. <https://doi.org/10.1101/615013>
- Watakabe, A., 2009. Comparative molecular neuroanatomy of mammalian neocortex: What can gene expression tell us about areas and layers? *Dev. Growth Differ.* 51, 343–354. <https://doi.org/10.1111/j.1440-169X.2008.01085.x>
- Werling, D.M., Pochareddy, S., Choi, J., An, J.Y., Sheppard, B., Peng, M., Li, Z., Dastmalchi, C., Santpere, G., Sousa, A.M.M., Tebbenkamp, A.T.N., Kaur, N., Gulden, F.O., Breen, M.S., Liang, L., Gilson, M.C., Zhao, X., Dong, S., Klei, L., Cicek, A.E., Buxbaum, J.D., Adle-Biassette, H., Thomas, J.L., Aldinger, K.A., O’Day, D.R., Glass, I.A., Zaitlen, N.A., Talkowski, M.E., Roeder, K., State, M.W., Devlin, B., Sanders, S.J., Sestan, N., 2020. Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic Coordination in the Developing Human Prefrontal Cortex. *Cell Rep.* 31. <https://doi.org/10.1016/j.celrep.2020.03.053>
- Wu, J., Zhao, Z., Shi, Y., He, M., 2022. Cortical VIP+ Interneurons in the Upper and Deeper Layers Are Transcriptionally Distinct. *J. Mol. Neurosci.* 72, 1779–1795. <https://doi.org/10.1007/s12031-022-02040-8>
- Yao, Z., van Velthoven, C.T.J., Nguyen, T.N., Goldy, J., Seden-Cortes, A.E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., Ding, S.L., Fong, O., Garren, E., Glandon, A., Gouwens, N.W., Gray, J., Graybuck, L.T., Hawrylycz, M.J., 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, S.M., Tieu, M., Torkelson, A., Tung, H., Ward, K., Dee, N., Smith, K.A., Tasic, B., Zeng, H., 2021. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* 184, 3222–3241.e26. <https://doi.org/10.1016/j.cell.2021.04.021>
- Yilmaz, M., Yalcin, E., Presumey, J., Aw, E., Ma, M., Whelan, C.W., Stevens, B., McCarroll, S.A., Carroll, M.C., 2021. Overexpression of schizophrenia susceptibility factor human complement C4A promotes excessive synaptic loss and behavioral changes in mice.

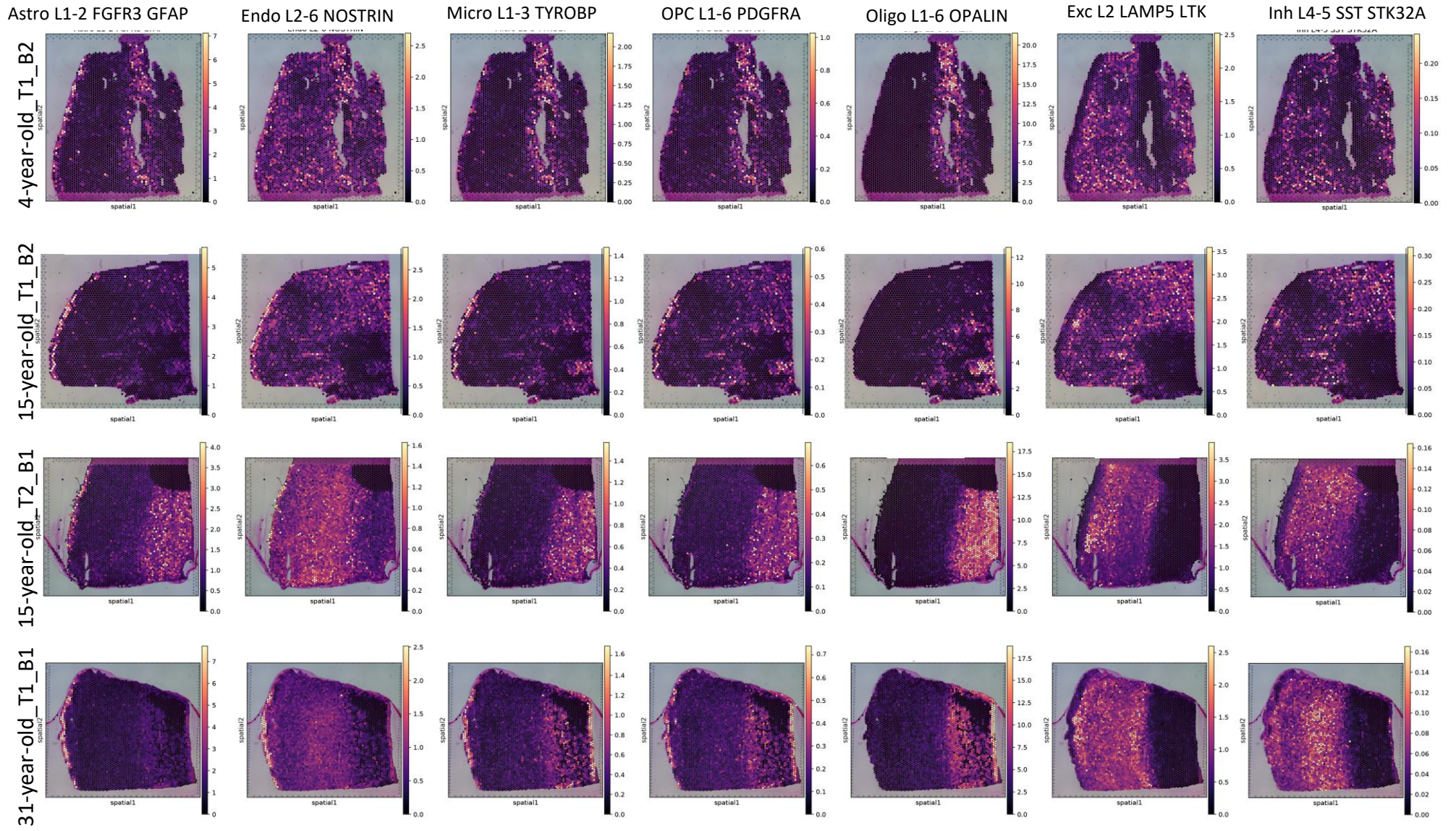
- Nat. Neurosci. 24, 214–224. <https://doi.org/10.1038/s41593-020-00763-8>
- Zhao, E., Stone, M.R., Ren, X., Guenthoer, J., Smythe, K.S., Pulliam, T., Williams, S.R., Uyttingco, C.R., Taylor, S.E.B., Nghiem, P., Bielas, J.H., Gottardo, R., 2021. Spatial transcriptomics at subspot resolution with BayesSpace. Nat. Biotechnol. 39, 1375–1384. <https://doi.org/10.1038/s41587-021-00935-2>
- Zhou, Y., Jia, E., Pan, M., Zhao, X., Ge, Q., 2020. Encoding method of single-cell spatial transcriptomics sequencing. Int. J. Biol. Sci. 16, 2663–2674. <https://doi.org/10.7150/ijbs.43887>
- Zhou, Z., Wang, S., Zhang, D., Jiang, X., Li, J., Gu, Y., Sun, H., 2020. Single-cell transcriptomic analysis identifies neocortical developmental differences between human and mouse. bioRxiv 2020.04.23.056390.
- Zhuang, P., Zhang, H., Welchko, R.M., Thompson, R.C., Xu, S., Turner, D.L., 2020. Combined microRNA and mRNA detection in mammalian retinas by in situ hybridization chain reaction. Sci. Rep. 1–10. <https://doi.org/10.1038/s41598-019-57194-0>



Supplementary Figure 1: Tapestation trace indicates the cDNA fragments size distribution after cDNA amplification for the 1st and 2nd Visium experiments. Lower and Upper markers used are included in each trace for each sample. The region that was set for the calculation of cDNA concentration is indicated by the vertical green lines on either side of the library trace.

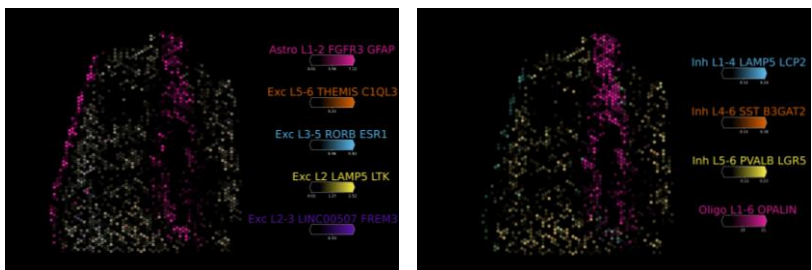
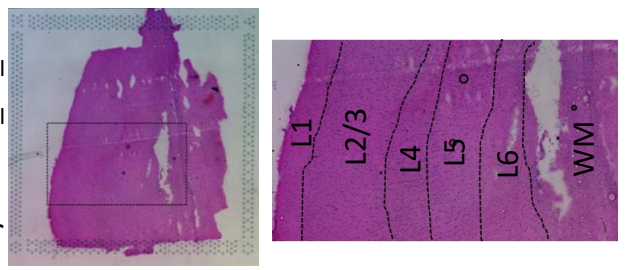


Supplementary Figure 2: Tapestation trace which shows the fragment size distribution after PCR Indexing for the 1st and 2nd Visium experiments. Lower and Upper markers used are included in each trace for each sample. The region that was set for the calculation of cDNA concentration is indicated by the vertical green lines on either side of the library trace.

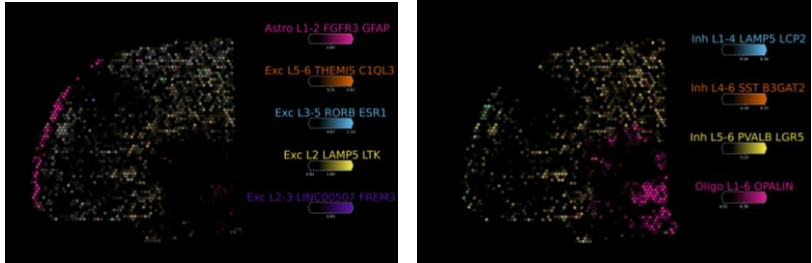
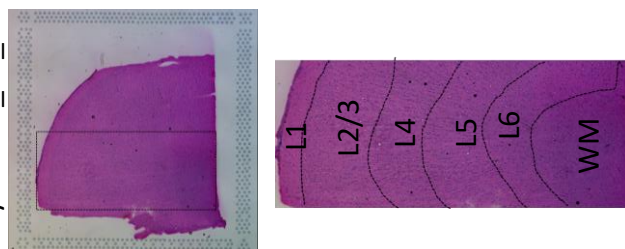


Supplementary Figure 3 : Spatial mapping of cell types in the human temporal cortex. Estimated cell type abundance plots for a selection of the major brain cell types in a representative section (technical replicate) for each patient sample. Brighter signal indicates a high abundance of the cell type.

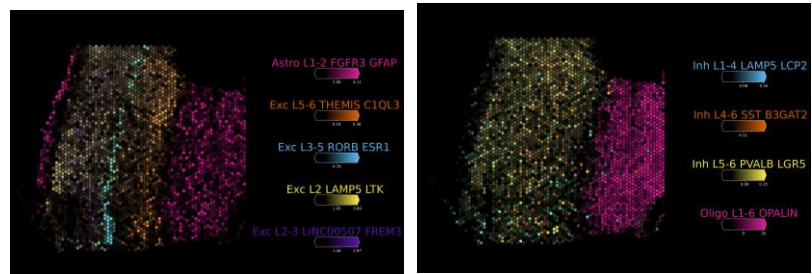
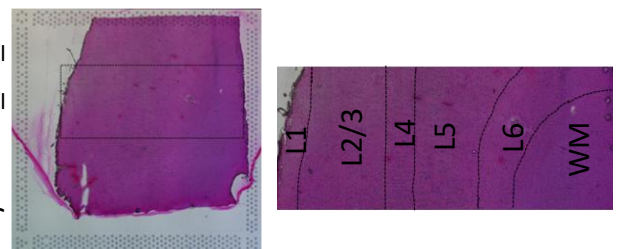
4-year-old_T1_B2



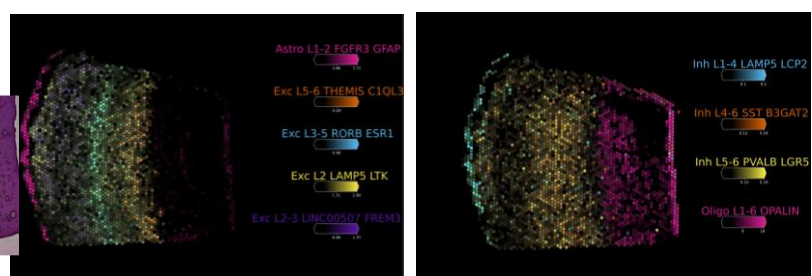
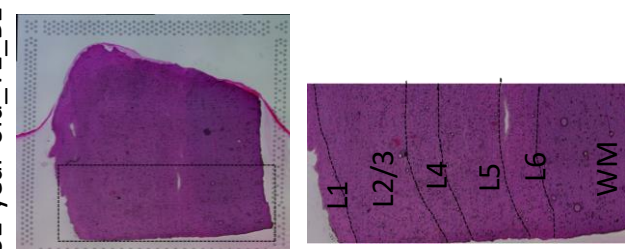
15-year-old_T1_B2



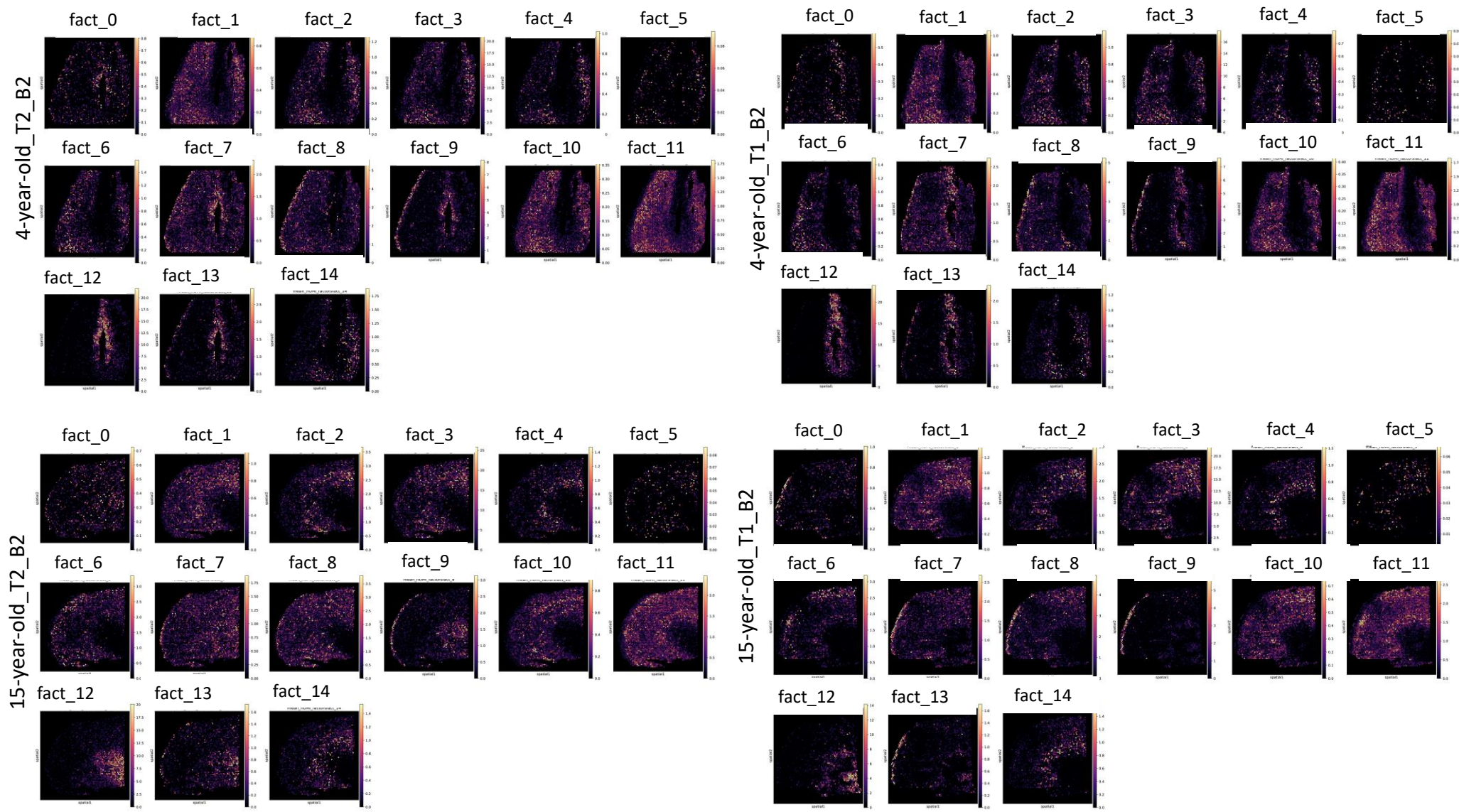
15-year-old_T2_B1



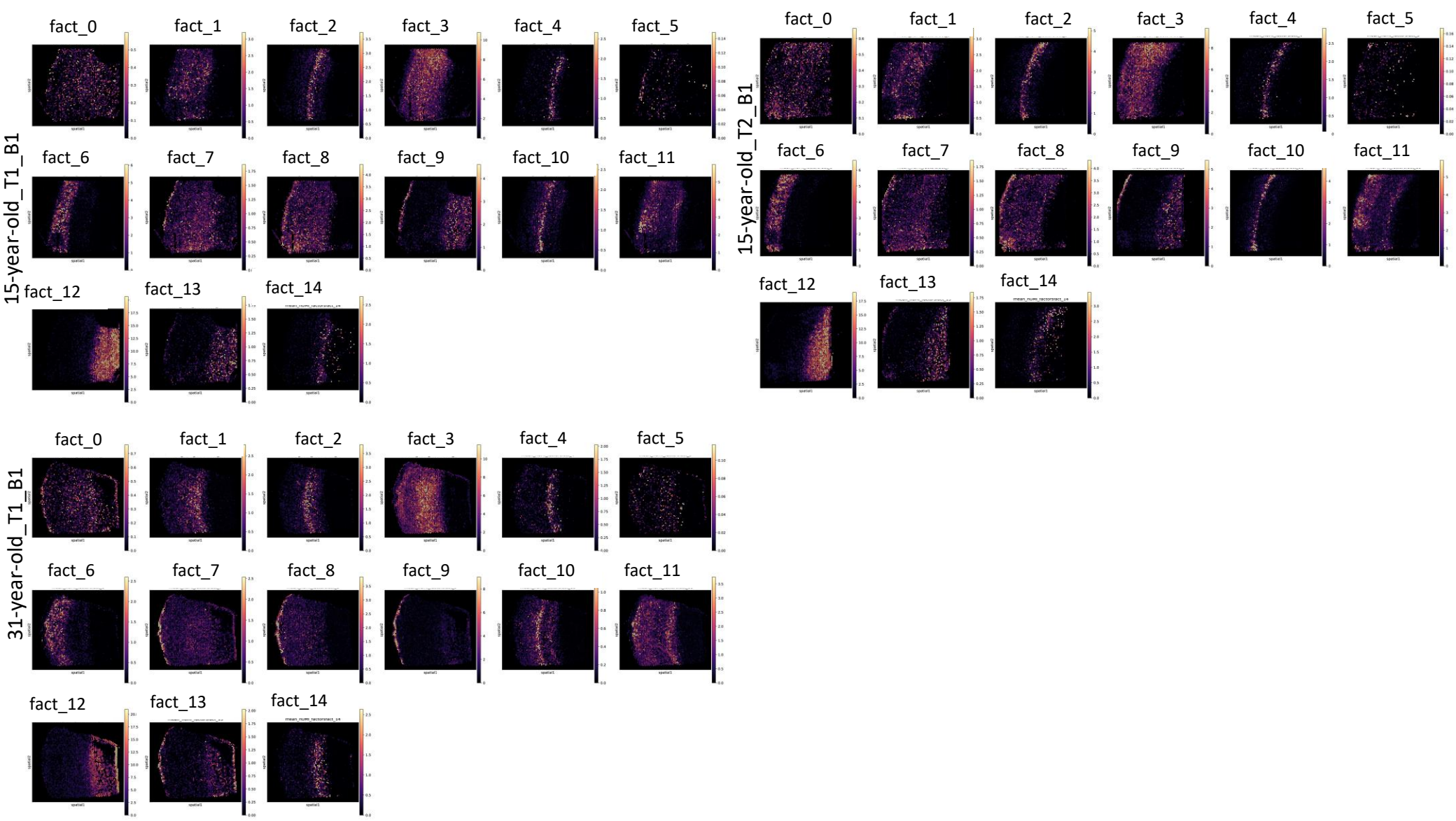
31-year-old_T1_B1



Supplementary Figure 4: Spatial mapping of layer-specific cell types.
A) H&E reference images for selected technical replicate temporal cortex sections generated during Visium Library preparation. The box indicates the region of the section that is magnified in the adjacent panel. B) Magnified H&E images annotated with the estimated boundaries of cortical layers 1-6 and the white matter (WM) marked by black dotted lines. C) Estimated cell type abundances (colour intensity) for a selection of cell types that clearly denote the cortex layers, especially in the 15-year-old & 31-year-old samples.

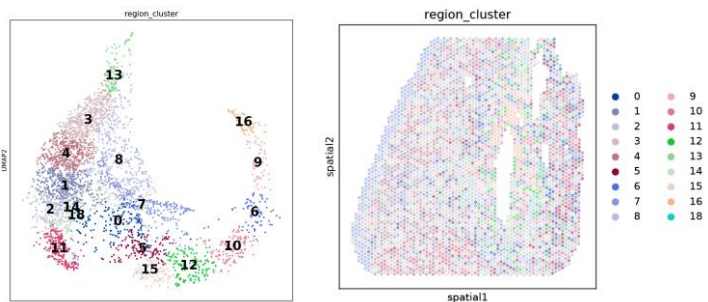


Supplementary Figure 5: Mapping the cellular compartments using non-negative matrix factorisation (NMF).

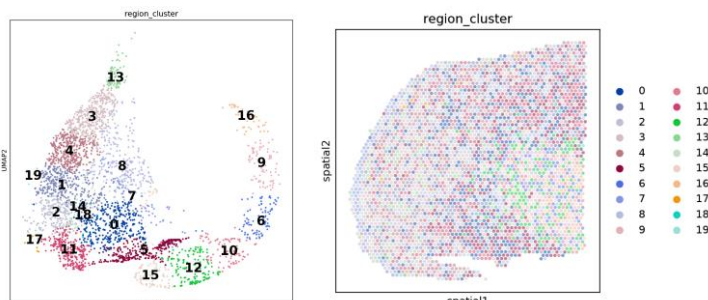


Supplementary Figure 6: Mapping the cellular compartments using non-negative matrix factorisation (NMF).

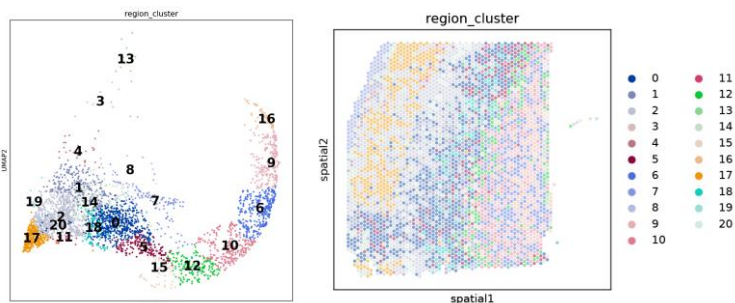
4-year-old_T2_B2



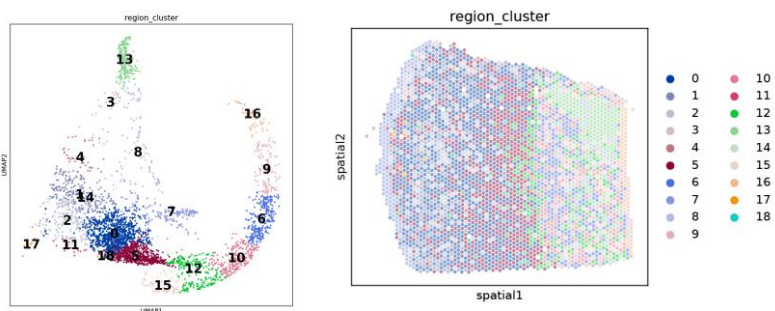
15-year-old_T2_B2



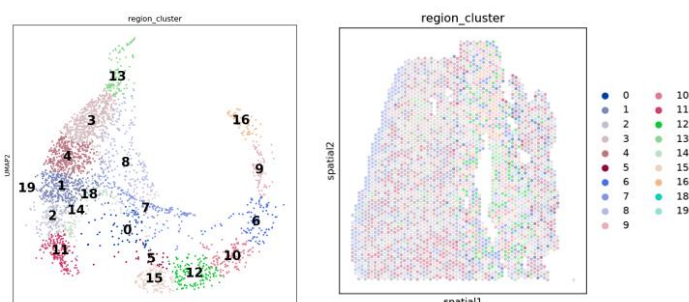
15-year-old_T1_B1



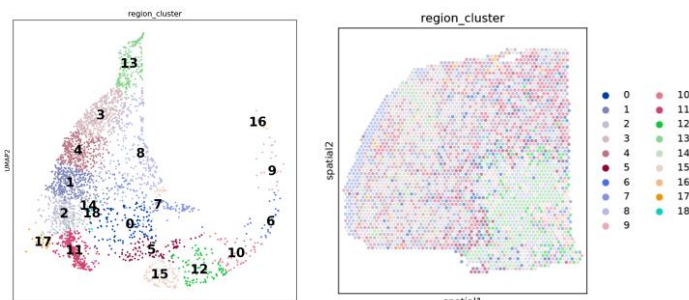
31-year-old_T1_B1



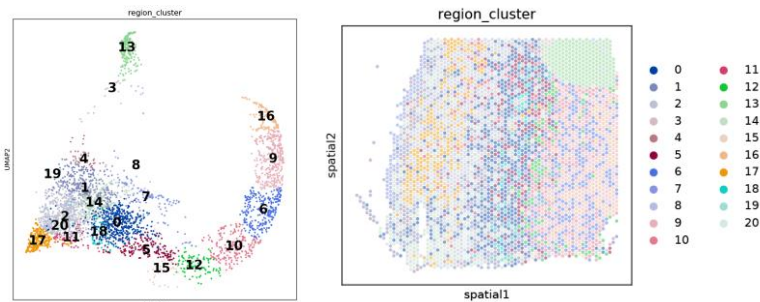
4-year-old_T1_B2



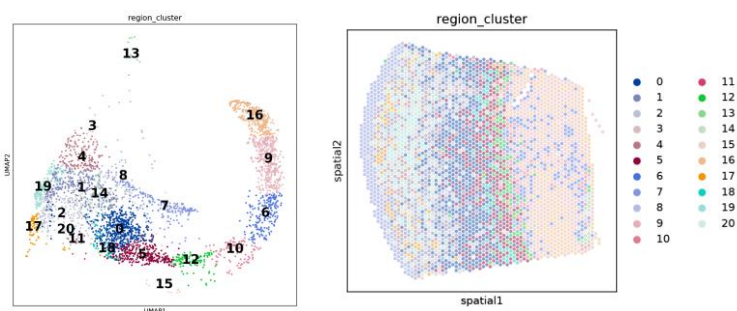
15-year-old_T1_B2



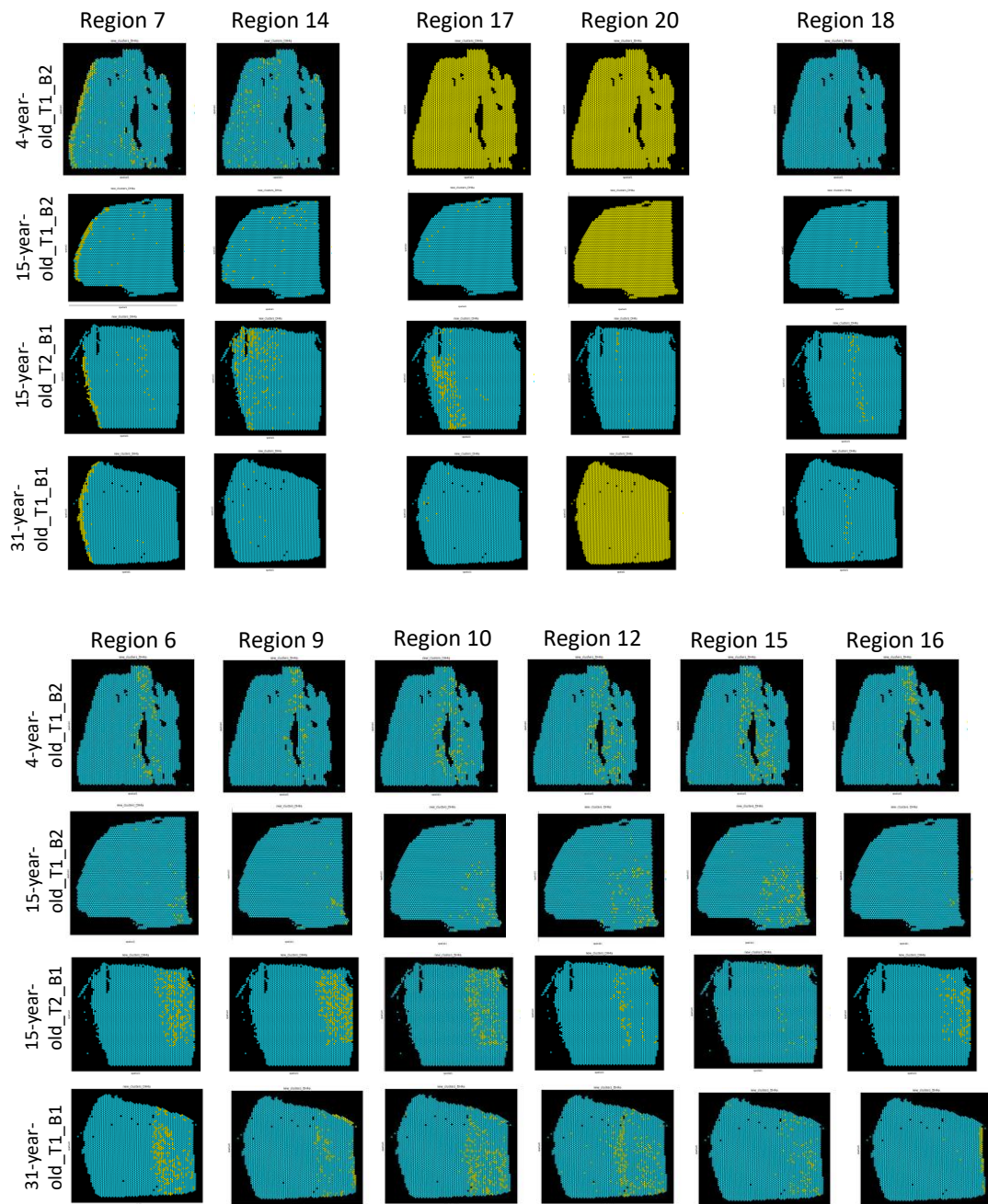
15-year-old_T2_B1



31-year-old_T2_B1



Supplementary Figure 7:
UMAP plots of the Visium
data coloured by Leiden
clusters for the individual
samples with their
corresponding location laid
on the tissue section image



Supplementary Figure 8: Spatial plots showing the location of individual Leiden clusters in yellow in representative samples (note: yellow colour across the entire section for Region 17 and 20 indicates the absence of any spots for that cluster)