High throughput molecular connectomics with barcoded rabies viruses

Alexander Becalick

University College London

Division of Biosciences

Submitted to University College London (UCL) in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy.

Primary supervisor: **Dr. Petr Znamenskiy**

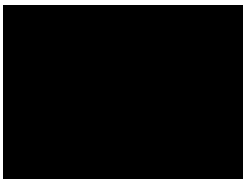
Secondary supervisor: **Prof. Kenneth Harris**

Word count: **49556**

Thesis submission date: **31/01/2025**

Declaration

I, **Alexander Becalick**, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



Alexander Becalick
London, United Kingdom
31/01/2025

Abstract

Neural circuits are comprised of diverse neuronal cell types with distinct connectivity motifs. Knowledge of this connectivity is crucial for furthering our understanding of circuit function. However, the throughput of conventional methodologies for studying the synaptic connectivity of individual neurons is limited. This prevents the ability to practically perform comparative connectomic studies or to systematically determine wiring patterns between the many different cell types. In this thesis, I describe the application of barcoded rabies viruses for mapping these connections in the mouse brain. Methods relying on DNA sequencing of oligonucleotide "barcodes" have previously been used to uniquely label large populations of neurons. To this end, I have generated libraries of rabies viruses containing barcode sequences with sufficient diversity to uniquely label populations of over 1000 starter cells in a single experiment. Taking advantage of the ability of rabies virus to spread transsynaptically between connected neurons, infected starter cells can then transmit their barcodes to their direct presynaptic partners. Presynaptic barcodes are then matched to their respective starter cells to reconstruct putative synaptic connections. To preserve information about the spatial location of barcoded cells and avoid cell loss resulting from tissue dissociation, I have used in situ sequencing to read out viral barcodes alongside endogenous transcripts of cell-type marker genes from sections of mouse brain. This has enabled me to relate the molecular identity of barcoded neurons with their connectivity. I then explore the feasibility of using this technique to determine the connectivity of neurons in the mouse visual cortex.

Impact Statement

Understanding the intricate network of neuronal connections in the brain, often referred to as the "connectome", is central to deciphering how information flows through neural circuits and how cognitive functions arise. However, mapping these connections across the large volumes of tissue that individual neurons can span, has posed significant challenges.

In this thesis, I have developed a high-throughput method for mapping the connectivity of large populations of neurons at once, by labelling them with unique "barcodes" that can be spread from neuron to neuron using barcoded rabies viruses. This approach provides single-cell granularity to the connectomic information which it produces, in contrast to established connectomic methods which have traditionally required sacrificing either throughput or cellular resolution. This offers the ability to study all manner of mammalian neural circuits for which rabies virus is capable of infecting, and will provide researchers with a new tool to bridge the gap between the complex network of neuronal connections and the distinct molecular identities of those neurons.

The methods described in this thesis represent a significant step toward a scalable connectomic platform: one that could be applied routinely to diverse brain regions and multiple animal models to study connectivity in a variety of contexts. This opens new avenues for comparative studies of neural circuits, both in normal physiology and potentially in disease states such as dementia or psychiatric disorders. Moreover, the development of a high-throughput spatial transcriptomics imaging and fluidics system that I have described in this thesis provides great versatility in its potential uses by bringing down the cost and increasing the scale at which gene expression can be profiled in biological tissue, at subcellular resolution.

Acknowledgements

This PhD has been a journey full of the most welcome surprises from start to end. I can wholeheartedly say that I would not have made it this far without the support and friendship of both those who have been in my life for years, and those for whom this PhD has given me the fortune of meeting more recently. I consider myself particularly lucky to have become surrounded by such good company.

First, I would like to thank Dr Petr Znamenskiy, for being a truly unrivaled mentor, a fountain of knowledge, and in some of my most despondent experimental moments, an unfaultering source of optimism.

I would like to thank the members of my thesis committee - Prof. Kenneth Harris, Dr. Andreas Schaefer, Dr Jonny Kohl and Dr. Samuel Rodriques for all of their feedback and advice throughout my PhD. A special thank you to Prof. Kenneth Harris for making me feel a welcome member of his own lab's community, in which I found much support and good humour.

I also would like to give a personal thank you to every Znamenskiy lab member. To Antonin, for being an excellent teacher and general miracle worker. Whilst he makes things look effortless sometimes, I've seen the amount of time and thought that he's put in to helping us all and I'm very grateful to him for it. To Yiran, for keeping me company at every milestone and making the whole trip better for it. To Benita, for many years of kindness and good advice, to Caroline for her thoughtfulness, genuine caring and an infectious enthusiasm for science and to Antonio for always finding a way to make me smile. I would like to thank the greater Neurophysiology community for all of these years of companionship, collaboration and comradery. Thank you to Gaia, Yuxin and Alex for giving me the best footsteps to follow in. To all my friends watching on from outside of the sphere of academia, I would like to thank them all for all their support. This experience has made me appreciate everyone all the more for it. To my brother Luke, and to my parents, I will always be grateful to such a loving family by my side. I can't thank them enough.

And finally, to the best companion of them all. Thank you Aiste.

Contents

Table of Acronyms

Notation	Description
ACAd	Dorsal part of anterior cingulate area.
ACAv	Ventral part of anterior cingulate area.
AIp	Posterior agranular insular area.
AId	Dorsal agranular insular area.
AIv	Ventral agranular insular area.
AUDd	Dorsal auditory area.
AUDp	Primary auditory area.
AUDpo	Posterior auditory area.
AUDv	Ventral auditory area.
ECT	Ectorhinal area.
FRP	Frontal pole, cerebral cortex.
GU	Gustatory area.
ILA	Infralimbic area.
MOp	Primary motor area.
MOs	Secondary motor area.
ORBl	Lateral orbital area.
ORBm	Medial orbital area.
ORBvl	Ventrolateral orbital area.
PERI	Perirhinal area.
PL	Prelimbic area.
RSPagl	Retrosplenial agranular area.
RSPd	Dorsal part of retrosplenial area.
RSPv	Ventral part of retrosplenial area.
SSs	Supplemental somatosensory area.
SSp-bfd	Primary somatosensory area, barrel field.
SSp-ll	Primary somatosensory area, lower limb region.
SSp-m	Primary somatosensory area, mouth region.
SSp-n	Primary somatosensory area, nose region.
SSp-tr	Primary somatosensory area, trunk region.
SSp-ul	Primary somatosensory area, upper limb region.

Notation	Description
SSp-un	Primary somatosensory area, unassigned.
TEa	Temporal association areas.
VISC	Visceral area.
VISa	Anterior visual area.
VISal	Anterolateral visual area.
VISam	Anteromedial visual area.
VISl	Lateral visual area.
VISli	Laterointermediate visual area.
VISp	Primary visual area.
VISpl	Posterolateral visual area.
VISpm	Posteromedial visual area.
VISpor	Postrhinal visual area.
VISrl	Rostrolateral visual area.

List of Figures

Chapter 1

Introduction

Each new insight into the nature of the brain's structure has been made possible with advancements in the techniques available to study the components of neural circuits. The challenge of studying connectivity across the multitude of scales that are present in the brains of complex organisms has remained a barrier to our understanding of these circuits. Adding to this challenge, these circuits are comprised of massively heterogeneous populations of neurons (Gouwens et al., 2020; Lee et al., 2023; Scala et al., 2021; Yao, van Velthoven, et al., 2023; Zeisel et al., 2015) that play distinct roles in neural computations (Economo et al., 2018; Li et al., 2015; Pfeffer et al., 2013; Pi et al., 2013). Determining the interactions between these neuronal cell types requires methods that enable the study of the local and long-range connectivity of these cells, as information transfer within the brain occurs at both local (Ko et al., 2011; Lefort et al., 2009) and global levels (Harris et al., 2019; Winnubst et al., 2019). To address this need for a scalable method to study neuronal connectivity, I will describe in this thesis the development and application of a novel tool for studying the connectivity of neural circuits in a manner that is agnostic to the distances between neurons, whilst capturing connectomic information at the resolution of individual, transcriptomically defined neurons. In this chapter, I will review the currently available methods for studying the connectivity of neurons. Then, I will focus specifically on work that has leveraged advances in high throughput sequencing techniques to determine neuronal projection patterns and connectivity.

1.1 Mapping neural circuits

Having access to the whole circuit diagram of a brain, its "connectome", enables testing hypotheses of neural computation within the bounds of the known circuit architecture. The first connectome to be described was that of the 1 mm long nematode worm *Caenorhabditis elegans*, which contains a total of 302 neurons (White et al., 1986). Since then, the far more complex brains of a 1st instar *Drosophila melanogaster* larva (Wind-

ing et al., 2023) and an adult *Drosophila melanogaster* brain have been mapped in their entirety (Dorkenwald et al., 2024; Schlegel et al., 2024). These connectomes have already been used to confirm hypothesised circuits that drive walking movements (Yang, Brezovec, et al., 2024), flight turns (Ros et al., 2024) and halting motions in adult *Drosophila* and have been used to generate computational models of the *Drosophila* visual system (Lappalainen et al., 2024) and antennal grooming circuit (Shiu et al., 2024) which can replicate experimentally determined neuronal responses and generate new hypotheses to be tested about circuit dynamics.

Producing a whole connectome is a process which contends with structures that span from the nanometer scale of synaptic connections, up to the macroscopic scale of the whole brain, which in mammals can mean tracing neuronal processes that travel centimetres between brain regions. Despite these challenges, work is ongoing to create larger and larger volumes of synaptically annotated connectomic data from electron microscopy studies, including recent efforts to reconstruct a cubic millimetre of the mouse visual cortex (MICrONS Consortium et al., 2021) and human temporal cortex (Shapson-Coe et al., 2024). Given that these connectomic datasets still represent the reconstruction of only 0.2% of the volume of the average 435 mm^3 adult mouse brain (Wang et al., 2020), other approaches are necessary to study connectivity at various scales. Some neuroanatomical techniques are suited to studying macroscopic and mesoscopic bulk connectivity between brain regions, whilst other techniques allow us to study the connectivity of specific subpopulations of neurons. Below, I will outline the advantages and shortcomings of several of these techniques.

Volume electron microscopy and X-ray tomography

All of the whole-brain connectomes to date (Bumbarger et al., 2013; Dorkenwald et al., 2024; Ryan et al., 2016; Schlegel et al., 2024; Verasztó et al., 2020; White et al., 1986; Winding et al., 2023), have been generated using volume electron microscopy (vEM) techniques. vEM encompasses a range of methods for performing electron microscopy throughout thick tissue volumes using transmission electron microscopy and scanning electron microscopy imaging modalities. Scanning electron microscopy enables in situ tissue slicing with microtomes (Denk and Horstmann, 2004), ion beam milling (Sergey et al., 2018) or array tomography of pre-cut sections (Schalek et al., 2012). Transmission electron microscopy relies on serial tissue sectioning onto grids, enabling tomographic reconstruction of thicker tissue sections (Soto et al., 1994) or array-based imaging of ultrathin sections (Phelps et al., 2021; Yin et al., 2020). vEM provides the ability to resolve the nanoscale structure of synaptic contacts and generate whole-cell reconstructions of circuits, using heavy metal staining techniques to provide contrast to the membranes of cellular structures (Hua et al., 2015). This provides the ability to study the detailed morphology of neurons and non-neuronal cells that are captured within these vEM volumes.

Electron microscopy does not provide the ability to read out the molecular profile of cells as is possible with the vast array of labelling molecules that are compatible with light microscopy. However, the morphology of vEM reconstructed cells (Elabbady et al., 2024) can be used to assign them to cell types that have been previously defined by combinations of morphological, electrophysiological and molecular approaches (Gouwens et al., 2020). The connectomic information from electron microscopy studies has even been co-registered with functional activity by performing two-photon calcium imaging followed by vEM reconstruction (Bock et al., 2011; Lee et al., 2016; MICrONS Consortium et al., 2021).

While electron microscopy has proven useful for producing whole brain connectomes of smaller organisms, the issue of scale remains when tackling the mammalian brain. As this technique relies upon perfect registration between many thin imaging volumes to reconstruct the processes of neurons, the larger the total volume, the more likely it is that imaging or tissue defects will prevent the full connectivity of neurons from being reconstructed. A non-continuous imaging region of just a handful of damaged tissue sections can render the connectivity between the two regions impossible to measure. Advancements have been made in automating the process of segmenting these vast petascale datasets and dealing with tissue warping and other artefacts (Dorkenwald et al., 2022; Mahalingam et al., 2022; Schmidt et al., 2024), but they still require concerted efforts of manual proofreading to reconstruct the processes of neurons, which will only further increase in scale with larger imaging volumes. In the intermediate volumes that it is currently possible to produce, there is another issue that presents itself. The limited dimensions of these datasets result in the truncation of neuronal processes at the edges of the reconstructed volumes. It is estimated that even at the cubic millimetre scale, about one-third of cells within the mouse MICrONS volume have truncated dendrites (Elabbady et al., 2024). This truncation leads to an incomplete view of the connectivity of these neurons, which prevents the study of long-range connectivity in these small volumes. This also affects the conclusions which can be drawn from measures of local connectivity, especially for neurons with more tortuous local processes.

An emerging field of 3D-imaging which can resolve biological structures in a non-destructive manner is that of X-ray imaging. X-rays can penetrate much further into tissue than electrons, enabling imaging without the need to section samples into many ultra-thin slices (Du and Jacobsen, 2018). Generating sufficient contrast in biological samples with traditional laboratory X-ray sources that have low coherence and radiation flux has limited their resolving power (Albers et al., 2024). By taking advantage of synchrotron facilities as highly coherent and intense X-ray sources whilst using staining techniques similar to those used for electron microscopy to increase contrast, tomographic imaging techniques have resolved millimetre-sized samples at micron-level resolution using propagation-based phase contrast tomography (Bosch et al., 2022). Holographic nano-tomography has reached an even higher resolution of 87 nm (Kuan et al., 2020). Whilst this is enough to reconstruct

dense cell bodies and trace large neurites within tissue blocks across large distances, these resolutions cannot reliably resolve thin axon collaterals and synapses. More recently, the use of ptychographic tomography with high radiation X-ray doses exceeding 10^{10} Gy has resulted in a resolution of 38 nm in 10 µm diameter samples, enabling a 60% synapse detection rate at 80% precision (Bosch et al., 2023). With upcoming improvements to synchrotron X-ray sources (Yabashi and Tanaka, 2017), X-ray optics (Murray et al., 2019) and sample staining (Lu et al., 2023), X-ray-based imaging techniques may overtake volume electron microscopy in the future as the preferred method for producing dense reconstructions of brain volumes. Presently, X-ray-based techniques can deliver high-speed imaging of large tissue volumes but are limited by their trade-off between volume and resolution, preventing them from yet performing large-scale connectomic studies.

Electrophysiological recording and CRACM

Another way of directly measuring the synaptic connectivity between neurons is to perform pairwise electrophysiological recordings. By patching two or more neurons, current can be injected into each cell in turn, whilst searching for a postsynaptic response from the other cells. This approach can reveal not only the connectivity between neurons but also the synaptic strength and plasticity of these connections (Holmgren et al., 2003; Lefort and Petersen, 2017; Reyes et al., 1998). Paired patch recordings have been used to establish general wiring rules, such as the over-representation of highly connected motifs and reciprocal connections compared to a purely random model of connectivity (Peng et al., 2024; Perin et al., 2011; Song et al., 2005). Further work has demonstrated the dependence of connection probability on the intersomatic distance between neurons (Campagnola et al., 2022; Levy and Reyes, 2012; Perin et al., 2011), and the discovery that excitatory connectivity probability and connection strength reflect the similarity of neuronal responses in L2/3 pyramidal neurons of the mouse visual cortex (Cossell et al., 2015; Ko et al., 2011, 2013).

These electrophysiological recordings can also be combined with the use of cell-filling biocytin stains to morphologically define cell types (Gouwens et al., 2019; Jiang et al., 2015; Lee et al., 2021). This can be extended further to incorporate the transcriptomic characterisation of patched neurons by extracting and processing the cytoplasm of a patched neuron using RT-PCR (Pfeffer et al., 2013) and more recently, scRNAseq (Cadwell et al., 2016; Fuzik et al., 2016; Gouwens et al., 2020; Scala et al., 2021). Whilst this combination of morphological, electrophysiological and transcriptomic characterisation can reveal a lot about cell-type specific local connectivity, the process of patch-clamping itself is slow. Patching neurons in tissue is time-consuming and this process is usually done in a manner that is blind to connectivity before the neurons are patched, so finding pairs of connected cells is not a rapid process. This has led to the development of semi-automated multipatch setups that can increase throughput and reduce the amount of manual labour

per cell recording (Kolb et al., 2019; Koos et al., 2021; Peng et al., 2019).

These setups are still limited to studying the connectivity of small numbers of cells at a time, so inferences must be made from collections of these small networks when studying connectivity motifs (Vegué et al., 2017). Connectomic studies are almost exclusively performed in acute tissue slices, removing sources of mechanical instability, improving the ability to visualise patchable neurons and providing a controlled environment to measure the responses of connected cells. This results in a similar set of caveats to those found in volume electron microscopy studies of connectivity. By slicing tissue into 300 µm thick sections, it was estimated that in cat visual cortex, 90% of excitatory axons and 40% of inhibitory axons within the volume would be truncated (Stepanyants et al., 2009). This results in a significant decrease in the expected connectivity rate of neurons compared to values derived from reconstructions of neuronal processes (Levy and Reyes, 2012). As with volume electron microscopy, this presents a challenge for studying long-range connectivity and complicates the inferred connectivity rates when comparing neuronal cell types with varying levels of local tortuosity to their processes (Campagnola et al., 2022).

Considering the low throughput of paired patch experiments, extensions to whole-cell recordings have been created, including the use of photoactivatable neurotransmitters such as “caged” glutamate, which is inert until pulsed with UV light - resulting in local release and stimulation of nearby neurons (Katz and Dalva, 1994). This can be paired with laser stimulation to quickly sweep across populations of presynaptic cells whilst monitoring a postsynaptic cell with whole-cell recording. Further advancements led to the development of an optogenetic approach termed ChR2-assisted circuit mapping, which unlike caged glutamate stimulation, is genetically targetable (CRACM) (Petreanu et al., 2007). In this technique, neurons of interest are genetically modified or virally transduced to express the light-gated ion channel channelrhodopsin (ChR2). By delivering pulses of light to spatially restricted regions of the tissue, it is possible to selectively activate presynaptic cells or even subcellular compartments of their axons or dendrites (Galloni et al., 2022; Petreanu et al., 2009), while monitoring postsynaptic responses in a patched neuron. By combining cell-type-specific promoters or Cre-driver lines with AAVs expressing ChR2, CRACM can be used to study connections from genetically defined cell types (Hooks et al., 2015; Wang et al., 2007). This comes with the caveat that light-stimulated neurons can activate intermediate neurons which can synapse onto the recorded neuron. Whilst response latency between mono and disynaptic input can differ enough to distinguish these in some circumstances (Geddes et al., 2016), circuit and cell-specific variability in these latencies means it is often not possible to determine polysynaptic from monosynaptic input (Cho et al., 2013). Polysynaptic responses can be abolished by using tetrodotoxin (TTX) to block voltage-gated sodium channels, preventing action potentials in intermediate neurons, whilst the ChR2 mediated depolarisation can be enhanced by applying the potassium channel blocker 4-Aminopyridine (4-AP) to allow vesicle release to still occur in directly light targeted cells (Petreanu et al., 2009).

Light microscopy and viral techniques

Whilst electron microscopy can resolve structures at a much higher resolution than light microscopy, an optical approach enables diverse labelling strategies for visualising and characterising neurons. Fluorescent markers can be used to selectively label cells, proteins or transcripts of interest in order to detect specific neuronal populations. For morphological reconstruction, the contrast of cell-filling markers and the dramatic reduction in data size compared to vEM reconstruction reduce the burden of image segmentation. This has allowed bulk fluorescence tracing methods to visualise axonal projections that extend into multiple target regions, thereby constructing long-range projection maps (Harris et al., 2019). Despite these strengths, conventional fluorescent tracing is limited by the resolution constraints of light microscopy. In regions containing dense clusters of identically labelled neurons, it becomes impossible to resolve individual cells or trace fine neuronal processes. To produce individual neuron reconstructions, single-colour labelling must therefore be sparsified, necessitating the imaging of hundreds of brains to create comprehensive datasets (Economo et al., 2016; Gao et al., 2022; Winnubst et al., 2019). One strategy to mitigate this issue involves increasing the number of fluorescent markers used per experiment. For example, Brainbow employs a combinatorial expression system of different fluorescent proteins, allowing individual neurons to be tagged with unique colour signatures (Livet et al., 2007). This method enables the differentiation of neurons within close proximity by using their ratios of fluorescent expression. However, the limited spectral bandwidth of available fluorophores and the inconsistency of their expression through the processes of a neuron still constrains the number of distinct colour combinations that can be achieved, and the density of neurons that can be uniquely labelled. Furthermore, whilst it is possible to label pre and post-synaptic protein complexes to distinguish the locations of synapses along these reconstructed neurons (Gao et al., 2019), the diffraction limit of conventional light microscopy precludes the reliable detection of synaptic contacts compared to close synaptic oppositions, meaning that determining synaptic connectivity between neurons purely by expressing fluorophores is not possible in this manner.

Expansion microscopy

Expansion microscopy (ExM) was developed to circumvent the resolution limits faced by standard light microscopy. In ExM, biological molecules within the specimen are anchored to an expandable polymer matrix by a process of chemical crosslinking. The tissue is then isotropically expanded by introducing a buffer (Chen et al., 2016). This increases the distance between the previously densely packed biological structures, enabling better resolving power using conventional light microscopy systems. Because ExM can preserve native protein conformation, it can be integrated with existing labelling techniques such as Brainbow (Chang et al., 2017; Li et al., 2021; Tillberg et al., 2016) or multi-channel immunostaining to detect excitatory and inhibitory synapses with Bas-

soon, Homer1, Gephyrin and other synaptic proteins (Chang et al., 2017; Gao et al., 2019). ExM benefits from considerably cheaper, faster and simpler imaging than is achievable with electron microscopy whilst retaining the molecular context of the tissue. This could potentially enable the connectivity of molecularly defined cell types to be resolved in large volumes. However, the technique's resolving power is still directly related to the extent of tissue expansion achieved. As the distance between presynaptic terminals and post-synaptic boutons is typically around 20-25 nm in rodents (Lucić et al., 2005; Schikorski and Stevens, 1997; Zuber et al., 2005) a linear expansion factor of 10-20 fold would be required to resolve these structures. Expansion factors of 4-fold were described for first-generation expansion techniques (Chen et al., 2016; Tillberg et al., 2016), but subsequent techniques have enabled 16-22 fold iterative expansion (Chang et al., 2017; Sarkar et al., 2022) and recently single step 10 fold expansion has been made possible, albeit with fragile gel properties (Truckenbrodt et al., 2018). Expanding a mouse brain with a 10-fold linear expansion factor would result in a 1000-fold increase in tissue volume with a final size of 10 x 7.5 x 15 cm. Even the largest currently available light-sheet microscope is not capable of covering this imaging volume (Vladimirov et al., 2024). A sectioning-based imaging technique could fulfil this role, either with pre-expansion of the whole brain followed by fMOST imaging (Winnubst et al., 2019) or with sectioning followed by expansion and conventional imaging. The data size will likely be an order of magnitude smaller than those of similar vEM volumes (Shen et al., 2020) but the same concerns apply to the large amounts of manual proofreading needed for confirming synaptic contacts/neuronal process tracing in these hypothetical optical connectome datasets, as they do with vEM datasets.

Transsynaptic viral strategies

Each of the previously described imaging approaches reconstruct circuit connectivity by resolving individual synaptic contacts and tracing neuronal processes from presynaptic boutons to axonal terminals. This mapping requires imaging at nanometer resolution across millimetre- to centimetre-scale volumes, generating large datasets that are labour-intensive to process and proofread. An alternative strategy for detecting synaptically connected neurons without having to capture every synaptic contact in the tissue is to exploit transsynaptic labelling techniques, which rely on labels that can spread between connected neurons in a circuit.

Some of the earliest described transsynaptic tracing experiments used tritiated proline, which can undergo transsynaptic transport after incorporation into proteins within neurons (Dräger, 1974; Grafstein, 1971). Similarly, radiolabelled tetanus toxin was demonstrated to have retrograde transsynaptic labelling properties (Büttner-Ennever et al., 1981). These radioactive tracers were later replaced with alternatives that could be detected with light microscopy techniques. One of these protein-based tracers was wheat

germ agglutinin (WGA), a plant lectin which can bind to glycoconjugates in neuronal membranes and travel both anterogradely and retrogradely, whilst also being transported transsynaptically (Itaya and Hoesen, 1982). Conjugated to horseradish peroxidase (HRP) (Gonatas et al., 1979), WGA-HRP allowed researchers to detect anterogradely synaptically connected neurons, albeit with limited efficiency. These conventional tracers share the problem of being diluted as they label each consecutive neuron.

A number of viruses are capable of transsynaptic spread and unlike conventional tracers, they can replicate in the neurons that they infect, enabling amplification of labelling in second-order infected neurons. The application of transsynaptic viruses for studying the connectivity of neural circuits was first described using rabies viruses (Kucera et al., 1985), followed by similar studies using Herpes simplex virus type 1 (HSV 1) (Ugolini et al., 1987), vesicular stomatitis virus (VSV) (Lundh et al., 1988) and pseudorabies virus (PRV) (Strack et al., 1989). HSV-1 rapidly spreads in a primarily anterograde transsynaptic manner but is also capable of retrograde spread (Ugolini et al., 1987). Pseudorabies virus likewise has a very fast rate of spread, and whilst the PRV-Becker strain is transported in both anterograde and retrograde directions, the PRV-Bartha strain and any PRV virus lacking gE and Us9 gene expression, acts as a retrograde transsynaptic virus (Card et al., 1998; Lyman et al., 2007). All rabies virus strains have been shown to spread in an exclusively retrograde transsynaptic manner (Kelly and Strick, 2000; Tang et al., 1999; Ugolini, 1995), with exceptions of anterograde transport in some peripheral neuron populations (Bauer et al., 2014; Potratz et al., 2020; Zampieri et al., 2014). By performing time course experiments where tissue is studied at short intervals after initial infection, the progress of synaptic spread from these viruses can be tracked through higher-order multisynaptic pathways (Astic et al., 1993; Ugolini et al., 1989).

Whilst these polysynaptic tracing experiments using wild-type viruses proved useful for early circuit mapping studies, the ambiguity over the exact order of synaptic spread confounded the accuracy of circuit information that could be resolved using this method. Additionally, the rapidity of native HSV-1, PRV and VSV infection results in cytotoxicity over short time courses (Archin and Atherton, 2002; Aston-Jones and Card, 2000; van den Pol et al., 2009). A breakthrough in the use of transsynaptic viral tools was made by deleting the glycoprotein gene of the rabies virus from its genome and instead supplying its expression in *trans* within selected populations of initially infected "starter" neurons (Wickersham et al., 2007). This ensures that the rabies virus is only capable of spreading to neurons which are directly presynaptic to these starter cells. Once in the presynaptic cells, the rabies is still capable of replication and amplification of any transgene labels it may express, but it cannot continue to spread to further neurons in a polysynaptic manner as the glycoprotein gene which is necessary for transsynaptic spread (Etessami et al., 2000), is not present in these presynaptic neurons. Replacement of the native rabies glycoprotein coat with that of envelope protein EnvA (derived from the subgroup A avian sarcoma and leukosis virus,) can render the EnvA pseudotyped rabies virus incapable

of initially infecting mammalian neurons, as the EnvA receptor protein (TVA) that is required for EnvA-mediated entry into the cell, is expressed exclusively in avian species (Young et al., 1993). This means that initial rabies infection can be specifically targeted to a population of neurons that are directed to express TVA (Wall et al., 2010; Wickersham et al., 2007). This has enabled the precise exploration of specific neural circuits by studying the presynaptic populations of neurons from defined starter cells using virally expressed fluorescent labels to mark their presynaptic populations (Miyamichi et al., 2011; Stepien et al., 2010; Sun et al., 2014, 2019; Weible et al., 2010; Weissbourd et al., 2014). If the transsynaptic spread is initiated from populations of starter neurons, the resulting presynaptic neurons cannot be traced back to individual starter cells due to the resolution limits of light microscopy, giving a bulk view of connectivity onto specified postsynaptic populations. Measuring single-cell resolution presynaptic connectivity has been achieved with transsynaptic rabies viruses by initiating infection from single starter cells (Marshel et al., 2010; Masaki et al., 2022; Rancz et al., 2011; Rompani et al., 2017; Rossi et al., 2020; Vélez-Fort et al., 2014; Wertz et al., 2015). However, studying connectivity from single neurons does reduce the throughput at which connectivity can be measured.

1.2 Barcoded connectomics

Anthony Zador first proposed the idea of transforming the problem of determining connectivity into a problem of sequencing (Zador et al., 2012). By leveraging advances in high throughput sequencing techniques that have made the readout of billions of nucleotides per day a routine procedure, finding ways to sequence the connectome is a promising concept. In this next section, I will outline the progress which has been made in sequencing-based approaches for studying neuronal projections and connectivity. A shared tenet of all of these approaches is to use the vast combinatorial space of nucleic acid sequences to provide a way to uniquely label individual neurons with nucleotide "barcodes". With each additional base that is added to a nucleotide barcode, the combination of possible unique sequences is multiplied by four, resulting in an exponential increase in the number of possible unique sequences as the barcode is increased in length. As Zador notes (Zador et al., 2012),

> "A barcode consisting of even 20 random nucleotides can uniquely label 4^{20} $\approx 10^{12}$ neurons, far more than the number of neurons ($<10^8$) in a mouse brain."

Since this sequencing proposal was made, the first method to take advantage of high-throughput sequencing was MAPseq (Kebschull, Garcia da Silva, Reid, et al., 2016). In this method, barcoded libraries of Sindbis viruses are used to infect neurons, with each viral particle containing a random nucleotide barcode. The viruses replicate in the neurons and produce many copies of their barcode sequence, which is transported along axons

using an engineered RNA binding protein to the regions that the neurons project to. By microdissecting the brain and extracting the barcodes from each tissue section, high-throughput bulk RNA sequencing can be used to recover single-neuron resolution projection maps covering all of the microdissected regions. This method has since been used to study projection patterns from the auditory cortex, visual cortex, ventral hippocampus, medial prefrontal cortex, tegmental reticular nucleus, medial entorhinal cortex, brainstem and olfactory bulb (Chen, Chen, et al., 2022; Gergues et al., 2020; Han et al., 2018; Hausmann et al., 2022; Mathis et al., 2021; Tsoi et al., 2022; Webb et al., 2022; Zeisler et al., 2023). It has also been demonstrated that multiple projection source regions can be distinguished within single brains, thus increasing the projection mapping throughput (Huang et al., 2020).

Using microdissection and bulk RNA-seq of each piece of tissue for MAPseq is a fast approach for studying projection patterns, but it does prevent the integration of other sources of information about the neurons that are being studied at the single-cell level. For instance, it is not possible to determine the individual transcriptomic nature of the projection neurons. This has since been addressed by a variety of other barcoding techniques. The first of these to be established was a direct continuation of MAPseq - named BARseq, using the same barcoded Sindbis viral libraries to determine the projection patterns of neurons. Unlike MAPseq, BARseq uses in situ sequencing to detect the barcode sequences in the injection site directly from intact tissue (Chen et al., 2019). As the tissue is not homogenised, in situ sequencing retains the spatial information of each barcoded neuron and allows the simultaneous spatial detection of endogenous genes in these neurons for transcriptomic cell typing (Chen, Fischer, et al., 2024; Sun et al., 2021). This technique has since been used for mapping projections from the primary motor cortex and olfactory bulb (Chen, Chen, et al., 2022; Muñoz-Castañeda et al., 2020). Whilst in situ sequencing is used to detect barcodes sequences in the somata of the source neurons, the projection patterns of BARseq are still determined by using the microdissection and bulk RNAseq detection method from MAPseq. An extension to BARseq uses in situ sequencing to also detect the barcode sequences found in the axons of the projecting neurons (Yuan et al., 2024). This resulted in the ability to resolve axonal projections to different layers in distant cortical regions from individual neurons. The detection sensitivity is not high enough to reconstruct full axonal branches but it does enable laminar projection information to be extracted. BARseq has also been combined with expansion microscopy to assist in the reconstruction of neurons in a distance-independent manner by using the detected barcode sequences as a unique marker for each cell, even when continuous axonal tracing is interrupted (Goodwin et al., 2022). Detection sensitivity in the axonal processes was again a limiting factor for achieving this goal.

Outside of BARseq, a number of other techniques have also used barcodes to determine the projection patterns of neurons. CONNECTID is one such approach which builds upon MAPseq by performing single-cell RNA sequencing on FACS-sorted barcoded neurons

from the injection site of a MAPseq experiment. This allows the Sindbis viral barcodes to be associated with scRNAseq gene expression profiles. This method can provide a more complete readout of endogenous gene expression than BARseq, at the expense of lower efficiency of cell detection due to cell loss during tissue dissociation and FACS sorting, and loss of high-resolution spatial information about the source cells. Three other barcoding approaches have used AAV2retro instead of Sindbis virus to study project populations. This engineered serotype of AAV can retrogradely label cells after uptake from axonal terminals (Tervo et al., 2016). In VECTORseq, pre-existing AAV2retro vectors were repurposed, using their expressed transgenes as a form of simple barcode label which can each be uniquely distinguished using single-cell RNA sequencing or FISH detection performed on infected cells from the target site (Cheung et al., 2021). MERGEseq and Projection-TAGs also use the same premise as VECTORseq but both use dedicated barcode sequences to label different neuronal projection populations (Xu et al., 2024; Yang, Liu, et al., 2024). Whilst barcoded Sindbis and AAV2retro experiments can both produce projection information, the properties of these viruses provide advantages and disadvantages to both techniques. Sindbis is a cytotoxic virus which kills neurons after days of initial infection (Kuramoto, 2019). Thus, experimental timing must be balanced to achieve the highest levels of barcode expression before cellular transcription is disrupted. AAV2retro is not cytotoxic and can maintain stable transgene expression for long periods without affecting the transcriptome of infected cells (Howard et al., 2008). However, the anterograde nature of Sindbis provides a complete description of the projection patterns from a given injection site, whilst using AAV2retro requires prior anatomical knowledge of a circuit to select the sites to label from using retrograde infections.

Transsynaptic viral labelling strategies

All of the previously mentioned barcoding techniques produce neuronal projection information at the single-cell level, but they do not provide information on the synaptic connectivity between cells, as they use viral vectors that are incapable of transsynaptic spread. An early attempt at determining synaptic connectivity with barcoded viruses was made using pseudorabies viruses (PRV) (Oyibo et al., 2018). In this implementation of connectivity mapping, “invader” barcoded PRV amplicons are packaged by helper PRV viruses, enabling them to spread transsynaptically. These invader barcodes are fused using a recombinase with “host” barcode amplicons which cannot spread transsynaptically, resulting in barcode pairs derived from synaptically connected neurons which can be sequenced to determine a matrix of connected neighbours. This method relies upon the use of PRV-Becker as a helper virus to transmit the barcoded amplicons between cells, and as PRV-Becker is not a monosynaptically restricted virus, this can result in polysynaptic barcode transmission. The possibility of polysynaptic spread means that the barcode pairs generated in each cell cannot be interpreted purely as direct connectivity between

pairs of neurons.

Measuring direct connectivity between neurons with barcoded transsynaptic viruses requires monosynaptically restricting the viral spread. One way to achieve this is to use glycoprotein-deleted rabies viruses and this is the approach that I have pursued during the course of my PhD. Several other studies have since described approaches to generate barcoded rabies virus libraries and apply them in different contexts. Saunders et al. described optimisations to the viral library packaging process and applied the resultant barcoded rabies viruses to cultured mouse neurons, using single-cell RNA sequencing to detect rabies barcodes and gene expression from infected cells (Saunders et al., 2021). Single-cell RNA sequencing has also been used to detect barcodes from rabies libraries applied to astrocytes *in vivo* (Clark et al., 2021) and from organotypic slice cultures of human cerebral cortex (Shin et al., 2024). In each of these studies, cells were dissociated and FACS sorted to retrieve infected neurons for sequencing. Saunders et al. observed that they were able to capture 10% of infected neurons from culture after dissociation and FACS sorting. Low capture efficiency of barcoded neurons after dissociation and FACS-associated losses were also noted in several other barcoded rabies studies (Lin et al., 2024; Zhang et al., 2024). Shin et al. developed a method for targeting rabies barcodes to the nuclear membrane of infected cells to enable the use of nuclear dissociation instead of whole-cell capture. The infected cell capture rate for this was not estimated, but in comparison to scRNAseq a moderate reduction in barcode detection sensitivity was observed (a median of 135 UMIs per cell, 120 UMIs per nucleus, $p<0.05$). Rabies barcodes have also been detected using microdissection of infected brain regions and targeted amplification RNA-seq in a similar approach to MAPseq (Lin et al., 2024), resulting in a measure of broad area-to-area connectivity. Most recently, Zhang et al. have described the application of barcoded rabies viruses *in vivo* in an effort to determine connectivity in mouse somatosensory cortex, using BARseq to detect rabies barcode sequences (Zhang et al., 2024). I discuss my findings in comparison with the results of these rabies barcoding experiments in Chapter 6, along with a discussion of differences in methodology and analysis.

1.3 BRISC: Barcoded Rabies In Situ Connectomics

In this thesis, I will describe the development of Barcoded Rabies In Situ Connectomics (BRISC), a strategy for mapping brain-wide synaptic connectivity with single-cell resolution, while simultaneously preserving spatial and molecular information of cells. The core concept is to infect a sparse population of starter cells with uniquely barcoded rabies viruses (Figure 1.1 A), and then allow these viruses to spread retrogradely to presynaptic partners (Figure 1.1 B). Because each viral particle carries a distinct genetic barcode, cells connected to the same starter neuron inherit that same barcode, allowing one to

infer synaptic connections by identifying which cells share identical barcodes (Figure 1.1 C). Crucially, BRISC avoids the need to dissociate or sort infected cells: instead, it uses in situ sequencing to read out barcodes directly from intact tissue sections (Figure 1.1 D). This preserves the spatial context of each neuron and enables parallel detection of endogenous gene transcripts (Figure 1.1 E).

1.4 Thesis overview

In Chapter 2, I describe the generation of barcoded libraries of rabies viruses and my quantification of the barcode diversity of these libraries. I demonstrate that the diversity of these libraries is sufficient for uniquely labelling thousands of starter neurons in a single experiment. **In Chapter 3**, I describe my efforts to scale up the process of in situ sequencing, which is required for viral barcode and endogenous gene detection. I describe the development of a custom microscope setup and a microfluidic system for performing in situ sequencing on large volumes of tissue. **In Chapter 4**, I describe the application of in situ sequencing to detect a panel of marker genes for which I create detection probes. I demonstrate that these markers can be used to classify neuronal cell types in the mouse visual cortex while retaining spatial information for each cell. **Finally, in Chapter 5**, I describe the application of the barcoded rabies library to study *in vivo* connectivity in the mouse visual cortex. I demonstrate the detection of rabies barcode sequences

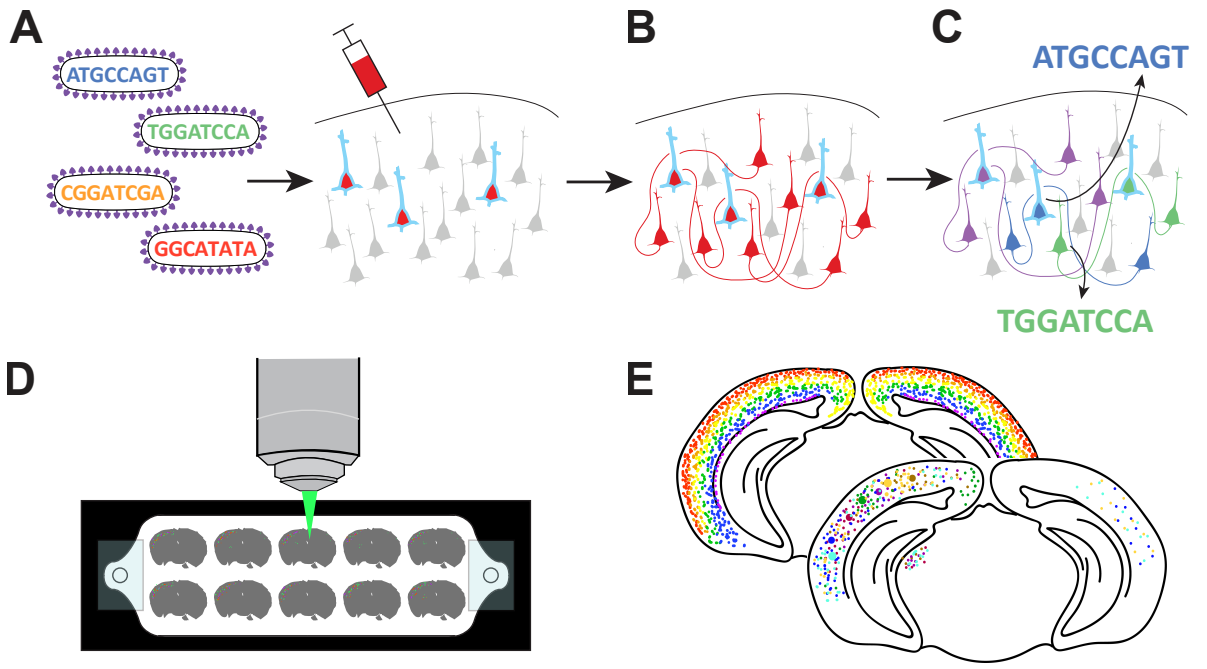


Figure 1.1: BRISC outline. (A) A barcoded library of rabies viruses (left) is injected into the brain, infecting a sparse population of starter cells (red with blue outline) **(B)** The barcoded rabies spreads transsynaptically to presynaptic partners **(C)** Presynaptic partners share the same barcode as their connected starter cell **(D)** In situ sequencing of tissue slices in a fluidic flowcell **(E)** Gene transcripts and viral barcodes are detected, maintaining spatial information.

from tissue and the assignment of these barcodes to individual cells which have been transcriptomically classified using the gene panel from Chapter 4. I analyse the resulting cell/barcode information and discuss remaining confounds for determining connectivity.

Chapter 2

Barcoded viral library production

2.1 Authorship declaration

I, Alexander Becalick, performed the majority of the experiments and analyses in this chapter, with the exceptions indicated here. Molly Strom of the Francis Crick Institute's Viral Vector Core performed cell culture, viral purification, and titering for the rabies packaging protocol.

2.2 Introduction

The ability to uniquely label cells with nucleotide barcode sequences has enabled the high-throughput study of lineage tracing (Naik et al., 2013; Schepers et al., 2008), clonal evolutionary fitness (Bhang et al., 2015; Levy et al., 2015) and neuronal projection patterns (Chen et al., 2019; Kebschull, Garcia da Silva, Reid, et al., 2016). The barcoded viral vector libraries used in these approaches make use of lentiviruses (Bhang et al., 2015; Levy et al., 2015; Naik et al., 2013; Schepers et al., 2008) and Sindbis virus (Chen et al., 2019; Kebschull, Garcia da Silva, Reid, et al., 2016), which are both amenable to viral rescue from plasmid genomes. Rabies viruses are well-established as a tool for studying synaptic connectivity, but generating barcoded viral libraries from rabies virus is not as simple as for other viruses, due to the low efficiency of viral rescue from their negative-strand RNA genome (Schnell et al., 1994). The generation of several barcoded rabies virus libraries has been described in recent years with varying levels of barcode diversity (Clark et al., 2021; Lin et al., 2024; Saunders et al., 2021; Shin et al., 2024; Zhang et al., 2024). In the following chapter, I will discuss the optimisations that I have implemented for barcoded plasmid generation and rabies virus packaging protocols to reduce the loss of barcode diversity. Using next-generation sequencing of the barcoded viral libraries, I will show the effects of viral packaging on barcode diversity and demonstrate that my resultant barcoded libraries have sufficient diversity for *in vivo* labelling experiments.

To effectively perform synaptic tracing of large populations of neurons with barcoded rabies viruses, it is essential to produce viral libraries with a high diversity of barcode sequences. In principle, each unique barcode sequence can be matched from labelled presynaptic neurons back to their originating starter cell which shares that sequence. However, the true source of a presynaptic cell labelled with a barcode that is present in more than one starter cell is ambiguous. Consequently, the probability of infecting more than one starter cell with the same barcode must be minimised. In a synaptic tracing experiment, the probability of initially infecting multiple starter cells with the same barcode is determined by the number of unique barcodes in the library, the abundance of each of those barcodes, and the number of times that the library is sampled in the experiment, e.g. the number of infection events. In the simplified case where the abundance of all barcodes in a library is even, then the fraction of uniquely labelled cells (F) can be represented by the following equation, where N is the total number of barcodes in the library and k is the number of infected cells (Kebschull, Garcia da Silva, Reid, et al., 2016):

$$F = \left(1 - \frac{1}{N}\right)^{k-1} \qquad \text{(Equation 2.2.1)}$$

During the production of barcoded viral libraries, barcoded plasmids are constructed and are then used to package functional viral particles. These processes of library generation result in an uneven distribution of barcode abundances in the final viral libraries. To account for this skewed barcode abundance distribution, we must individually weight the contribution of each barcode by the probability of selecting it from the library (Kebschull and Zador, 2018), so that for barcode $i = 1 \ldots N$ with a probability of selection p_i, the fraction of uniquely labelled cells (F) is defined as:

$$F = 1 - \sum_{i=1}^{N} p_i \left(1 - (1 - p_i)^{k-1}\right) \qquad \text{(Equation 2.2.2)}$$

Thus, in practice, it is important to preserve an even distribution of barcode abundance during library production whilst also generating a large number of unique barcode sequences to maximise the fraction of cells that can be unambiguously labelled in an experiment. The potential for barcode diversity loss is due to several factors. When generating barcoded genomic constructs, incorporating oligonucleotides with a region of random bases into a viral genome plasmid creates a plasmid library which is theoretically limited in diversity only by the randomness of the synthesised oligonucleotides (Saunders et al., 2021). However, this initial barcoded library of open-circular plasmids must then be amplified and supercoiled before it can be efficiently transfected into mammalian cells (Cherng et al., 1999; Remaut et al., 2006). The process of transformation into *E. coli* and the amplification that occurs in bacterial culture presents the first opportunity for barcode loss and competition within the library. Bacterial transformants can replicate at different rates within liquid culture, leading to a skew in barcode abundance distribution

(Joung et al., 2017). This can be limited by amplifying transformed libraries on a solid culture medium, which minimises the differences in replication rates between clones by limiting the local nutrient supply of each colony (Uzonyi et al., 2022).

Viral packaging from barcoded plasmids presents a second opportunity for skewing the abundance distribution of a barcode library. As a negative-strand RNA virus, rabies has a low rescue efficiency from cDNA. Whilst the genomes of positive-strand viruses can directly function as mRNA, negative-strand viruses need to form ribonucleoprotein (RNP) complexes to replicate and this requires viral proteins to initially be supplemented in trans (Roberts and Rose, 1998). Direct rescue of negative strand RNA viruses from negative strand genomes (gRNA) has repeatedly been unsuccessful (Conzelmann, 1996; Lawson et al., 1995; Radecke et al., 1995; Whelan et al., 1995) due to hybridisation between uncoated negative gRNA and positive-strand viral mRNAs, preventing RNP assembly (Schnell et al., 1994). Instead, rescue methods are mediated by transcribing an intermediate positive strand anti-genome agRNA from the plasmid, which forms RNPs unimpeded (Conzelmann, 2004). This rescue from an intermediate agRNA means that the initial assembly and packaging of viral particles must be mediated by supplying a minimum required complex of rabies genes in trans, as the anti-genome strand cannot be used to directly produce viral gene transcripts. Furthermore, transcription of the rabies virus anti-genomic RNA from the plasmid only produces a naked agRNA strand. Naked rabies agRNA cannot initiate replication or transcription on its own in the cell (Conzelmann, 2004) as it lacks the essential nucleoprotein (N) that mediates encapsidation of the RNA to form a nucleocapsid scaffold. This scaffold can be recognised by the viral polymerase complex of the rabies phosphoprotein (P) and large polymerase gene (L) (Albertini et al., 2011) which can then initiate transcription of all viral gene RNAs by stop/start transcription (Finke et al., 2000), and replication by transcription of the rabies negative-strand genome. This encapsidation step is a bottleneck of viral rescue.

In the first "reverse genetics" experiments carried out with rabies viruses, rescue was mediated from a 3980 bp minimal artificial rabies genome analogue, SDI-blue, which resulted in rescuing one out of 1000 cells (Conzelmann and Schnell, 1994). Using the full-length 12 kb transcript resulted in only one rescue event in 2 x 10^7 cells (Schnell et al., 1994). Increasing the efficiency of these viral rescue events will increase the number of unique barcodes present in a rabies viral library. It has been shown that the precise sequence of the 5' and 3' ends of rhabdovirus genomes is of critical importance to viral replication. An exact 3' terminus is required for replication in the closely related vesicular stomatitis virus (Pattnaik et al., 1992). Whilst the first rabies rescue constructs were designed with a hepatitis delta virus antigenomic ribozyme (HDVagrz) inserted at the 3' end of the rabies antigenome to generate an exact 3' terminus (Schnell et al., 1994), the HDVagrz core sequence that was used in those constructs has been shown to lack 5 nucleotides of the native ribozyme sequence, reducing its autocatalytic cleavage efficiency (Wadkins et al., 1999). Inserting these extra 5 nucleotides into rabies rescue constructs

has been shown to increase the efficiency of rabies rescue 10-fold (Ghanem et al., 2012).

Furthermore, to reach titres suitable for *in vivo* usage, each virus from the limited number of initial rescue events must be amplified through viral replication. Standard protocols for generating high-titre rabies virus stocks achieve this through multiple, prolonged stages of amplification to maximise the amount of virus produced (Osakada and Callaway, 2013; Wickersham and Sullivan, 2015). This long amplification process provides an opportunity for the amount of clonal expansion to vary between different barcoded virus-producing cells, introducing potential biases to the distribution of barcode abundances within the library. Several optimised protocols for barcoded rabies library production have circumvented this by increasing the scale of initial rescue and shortening the amplification stages (Saunders et al., 2021; Shin et al., 2024).

During rabies rescue, a fourth rabies gene must be supplied in trans: the rabies glycoprotein (G). This gene is not necessary for transcription or replication but is required for viral propagation. It is a type I membrane protein which interacts with cellular receptors to enable membrane fusion and viral entry to occur (Etessami et al., 2000). During packaging, this allows newly budded viral particles to infect further cells, amplifying the titre of the virus in this first stage. This glycoprotein was originally expressed in packaging cells by using a glycoprotein plasmid (Conzelmann and Schnell, 1994; Inoue et al., 2003), but stable cell lines that express the glycoprotein have since been produced (Osakada and Callaway, 2013; Sumser et al., 2022; Wickersham et al., 2010). For use in barcode labelling experiments, it is necessary to restrict the infectivity of the final barcoded viruses so that they can be targeted *in vivo* to infect a defined population of starter cells. This can be achieved by transferring glycoprotein-coated rabies from the supernatant of the initial glycoprotein-expressing rescue culture to a population of pseudotyping cells which express the envelope protein (EnvA) from avian sarcoma leucosis virus subgroup A (Young et al., 1993). After infection of these EnvA-expressing cells, rabies RNPs within the cells are coated with this alternative envelope protein upon budding, to produce pseudotyped EnvA-coated viruses (Wall et al., 2010; Wickersham et al., 2007). This step of reinfecting new cells with the supernatant from the first rescue transfection presents another potential bottleneck in maintaining the barcode diversity of the libraries due to barcode loss during infection and varying rates of viral replication across the population of successfully infected pseudotyping cells.

There are many strains of rabies virus, each with varying levels of cytotoxicity and neurotropism (Rasalingam et al., 2005; Tsiang et al., 1983). The Challenge Virus Strain (CVS) was originally isolated by Louis Pasteur from a rabid cow before being repeatedly passaged in rabbit and mouse brains (Habel, 1940; Pasteur, 1884; Sacramento et al., 1992). These passages led to mutations from the original ”street strain” into a ”fixed strain” which is adapted to the central nervous system (Lépine, 1938). The mouse-adapted CVS-24 strain was further passaged in BHK cells, resulting in the highly neurotropic CVS-

N2c strain (Morimoto et al., 1998). I used the CVS-N2c strain of rabies virus as the basis for my barcoded viral libraries due to its reduced cytotoxicity and enhanced transsynaptic neuronal transfer compared to the commonly used attenuated vaccine strain SAD-B19 (Reardon et al., 2016).

2.3 Methods

2.3.1 Viral library production

Libraries of barcoded rabies virus were generated in a three-stage process. Firstly, random barcode sequences were introduced into a rescue plasmid containing the rabies genome. These plasmids were then amplified and supercoiled during bacterial transformation. Finally, the barcoded genome plasmids were transfected along with rabies gene packaging plasmids into mammalian cell lines to produce barcoded viral particles.

Full length ribozyme addition

A CVS-N2c-ΔG-mCherry plasmid (Addgene plasmid # 73464, a gift from Dr Andrew Murray) was used as the basis for producing barcoded rabies libraries. The extended 3' sequence that is necessary to generate the super cutter variant of the hepatitis delta virus antigenomic ribozyme (HDVagrz) was introduced into the plasmid using a Q5 site-directed mutagenesis kit (New England Biolabs, E0554S). An oligonucleotide was inserted containing a sequence complementary to the 5' region of the hepatitis delta virus antigenomic ribozyme, along with 5 extra bases (ACGGA). Plasmids were extracted from resultant transformed *E.coli* and were Sanger sequenced using Supercutter_F2 and Supercutter_R forward and reverse primers to select a plasmid with the successful incorporation of the extra bases.

Barcode incorporation

An inverse PCR reaction was then performed on the super cutter plasmid to introduce the barcode sequences, using a modified version of a previous protocol for barcoding plasmids (Saunders et al., 2021). The barcode cassette was targeted downstream of the mCherry transgene by a pair of forward and reverse primers (Barcode_F and Barcode_R, CAB0001, CAB0002 for 20bp barcode). Both primers are comprised of a 5' tail containing six protective bases flanking an EagI restriction site, the forward primer also contains 20 bp of random nucleotides forming the barcode, with both primers containing a genome complementary 3' sequence. To maximise the amount of linear barcoded plasmid DNA that was produced for downstream steps, 96 PCR reactions were set up filling a 96-well PCR plate at 25 µl per well. This mix was comprised of 1250 µl Q5 Hot Start High-Fidelity

2X Master Mix, 125 µl forward primer (10 µM), 125 µl reverse primer (10 µM), 500 µl template plasmid (supercutter-CVS-N2c-ΔG-mCherry, 0.4 ng/µl) and 500 µl molecular grade water.

The following thermocycling conditions were used to generate the linear plasmid constructs with EagI 5' and 3' ends: 98 °C - 30 s, 35 cycles of 98 °C for 10 s, 72 °C for 30 s, 72 °C for 450 s, followed by 72 °C for 2 mins and 4 °C hold. After this barcode addition PCR, all of the wells were combined and the total volume of PCR product was run on a 1 % agarose TAE gel (Thermo Fisher Scientific, 16500500) at 120 V for 1 hour. Multiple columns of a Wizard SV Gel and PCR Clean-Up kit (Promega, A9281) were used to extract the PCR product from the gel, loading a maximum of 300 µg per column. The product was pooled and DNA concentration was determined with a Nanodrop 2000 spectrometer (Thermo Fisher Scientific). Secondary purification was then carried out with DNA Clean & Concentrator-5 columns (Zymo Research, D4004). Typically around 20 µg of linear DNA was recovered from a single 96-well plate.

The purified PCR product was then digested to expose sticky ends for ligation with the following reaction mix: 0.5 µl EagI-HF (New England Biolabs, R3505L), 0.25 µl DpnI (New England Biolabs, R0176L), 500 ng linear barcoded DNA, 5 µl CutSmart buffer (New England Biolabs, B6004S) and molecular-grade water up to 50 µl. The total amount of purified PCR product was variable, so the volumes of this reaction mix were scaled based on the measured yield of linear DNA. Digestion was carried out at 37 °C for 1 hour and then heat inactivation was carried out at 80 °C for 20 minutes. The sticky-ended products were then ligated into open-circular plasmids by directly adding T4 ligase (New England Biolabs, M0202T) and ATP (Thermo Fisher Scientific, PV3227) to the reaction mix at a ratio of 0.1 µl ligase and 5 µl ATP per 44.9 µl of sticky-ended product. This was incubated at 4 °C for 2 hours and then heat inactivation was carried out at 65 °C for 20 minutes.

To remove leftover linear DNA a final Exonuclease V (New England Biolabs, M0345S) digest was carried out by adding to the reaction mix in a ratio of 1 µl ExoV, 1.4 µl NEBuffer4 (New England Biolabs, B7004S) and 6.4 µl ATP per 55.5 µl of ligated product. Digestion was carried out at 37 °C for 1 hour and then heat inactivation was carried out at 70 °C for 30 minutes. The resulting product was purified using DNA Clean & Concentrator-5 columns and then pooled before measuring the final concentration with a Nanodrop 2000 spectrometer.

Barcoded plasmid transformation

Next, plasmids were transformed into MegaX DH10B T1R Electrocompetent cells. (Thermo Fisher Scientific, C640003) to preserve the diversity of the open circular plasmids whilst amplifying and supercoiling them for mammalian transfection. Six electroporation cuvettes (BIO-RAD, 1652082) were chilled on ice, then 40 µl of electrocompetent cells were

mixed with 400 ng of circularised DNA per cuvette and added. Samples were immediately electroporated at 2.0 kV, 200 Ω, 25 µF using a GenePulser XCell (BIO-RAD, 1652662). 1 ml of pre-warmed SOC media was then added to each cuvette and placed in a round-bottomed tube in a shaking incubator for 1 hour at 37 °C. All samples were then pooled, diluted to 20 ml with LB broth and then 2 ml was spread evenly across each of ten prewarmed LB Carbenicillin (245 mm Square BioAssay plates (Corning, 431111) with inoculating L-shaped spreaders (Thermo Fisher Scientific, 11836191)).

LB Carbenicillin plates were prepared freshly two days prior to transformation by mixing 200 ml LB agar with 200 µl Carbenicillin stock solution (100mg/ml) per plate. The plates were incubated overnight at 30 °C and then all colonies were scraped from the plates, adding 20 ml LB to wash and collect the colonies. The bacteria were pooled and centrifuged at 5000 g for 10 minutes to collect a pellet which was weighed. Separate ZymoPURE II maxiprep columns (Zymo Research, D4202) were used to extract endotoxin-free plasmid DNA from 0.5 g of pellet each.

2.3.2 Rabies packaging and rescue

The following packaging protocol was carried out by Molly Strom (Francis Crick Institute viral vector core). HEK293T cells were transduced with an EF1a-N2cG-puromycin lentiviral vector to produce a stable cell line expressing the N2c-glycoprotein (HEK293TG). Rabies viruses were rescued from the barcoded genome plasmid by transfecting HEK293TG cells with the genome plasmid along with plasmids containing the SADB19 rabies genes N, P, L and G as well as T7 polymerase in the following protocol:

Day 0: 6-well dishes were seeded with HEK293TG cells at a density of 6 x 10^5 cells/well in DMEM - 5% FBS. Day 1: The following transfection mix was prepared - 6 µg pTIT-N, 3 µg pTITL, 5 µg pTITP, 6 µg pTIT-G, 8.5 µg pCAGGS T7 and 12 µg Barcoded library in 430 µl of Opti-MEM. Then, 85 µl of PEI MAX (1 mg/ml) was added to the mix and incubated at room temperature for 20 minutes. 80 µl of this was added to each well. Day 2: The media was changed with fresh DMEM - 10% FBS. Day 4: The cells were split in DMEM - 10% FBS at a ratio of four wells to a T175 flask. Day 7/8: When cells reached 70% confluency they were transfected with further glycoprotein plasmid in the following transfection mix: 13 µg CMV-RG, 1 ml Opti-MEM, 40 µl of PEI MAX (1 mg/ml). This was added to the cells and incubated for 6-8 hours before replacing with fresh DMEM - 10% FBS media. Day 10/11: Supernatant was collected and filtered through 0.45 µg filters. The supernatant was titered by serial dilution onto HEK293TG cells in 6-well plates, with a typical titer of 1 x 10^7.

This supernatant was used to infect BHK cells stably expressing the EnvA coat protein for pseudotyping: Day 0: 150 mm dishes were seeded at a density of 1 x 10^7 cells per dish with BHK-EnvA cells in DMEM - 10% FBS. Day 1: Enough supernatant was added

to each dish to achieve an MOI of 0.3. The cells were incubated with supernatant for 6-8 hours and then washed with PBS four times, then split and replated in DMEM - 5% FBS at a ratio of 1:3. Day 3: Supernatant was collected and filtered through 0.45 µg filters and fresh DMEM - 10% FBS was added to the plates. The supernatant was centrifuged at 70,000 g in an SW32.1 rotor for 2 hours at 4 °C and then the viral pellets were resuspended in approximately 250 µl PBS. Day 4: Day 3 was repeated and then the remaining dishes were disinfected and disposed of. The titer of the concentrated viral solutions was determined by applying serial dilutions of the virus to HEK293T-EnvA and HEK293T cells. The counts of mCherry+ HEK293T-EnvA cells determined the pseudotyped rabies viral titer whilst the HEK293T mCherry+ count determined the titer of glycoprotein-coated viral contamination.

2.3.3 Sequencing the barcode libraries

Viral RNA extraction and reverse transcription

The diversity of the barcoded libraries was determined by counting the abundance of rabies plasmid or viral barcodes with a UMI-tagging approach. For viral libraries, rabies virus was sampled at various time points during viral packaging from PEG-precipitated cell culture supernatant or by sampling ultracentrifuge concentrated final viral aliquots. Viral genomic RNA was extracted using the Quick DNA/RNA viral kit (Zymo Research, D7020) and eluted into 15 µl of RNase-free water. Purified viral barcode RNA was then reverse transcribed using Superscript IV reverse transcriptase (Thermo Fisher Scientific, 18090050) at 50 °C for 10 minutes with 1 µl of 2 µM UMI_P5_Trueseq primer, before heat inactivation at 80 °C for 10 minutes. RNA was subsequently digested with RNase H (New England Biolabs, M0297S) at 37C for 20 minutes. The primer used for reverse transcription (UMI_P5_Trueseq) contained an overhang of 22 bp P5 Trueseq for nested adaptor PCR, 12 bp of random UMI sequence and 20 bp of rabies homologous sequence for binding upstream of the barcode.

Barcoded plasmid amplicon generation

This same UMI_P5_Trueseq primer was used for producing UMI tagged amplicons from the barcoded plasmid libraries in a two-cycle limited PCR reaction. 5 ng barcoded plasmid was added to 16 µl Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, M0494S) with 1.25 µl each of UMI_P5_Trueseq and P7_universal_reverse primers and PCR thermocycled with the following conditions: 98 °C - 3 mins, two cycles of 98 °C - 30 s, 67 °C - 20 s, 72 °C - 30 s, followed by 72 °C - 2 mins and 12 °C hold.

Viral/plasmid amplicon indexing

The resultant cDNA from the viral samples or amplicons from barcoded plasmids was size selected and purified with KAPA Pure Beads (Roche, KK8000) in a 1:1 ratio, eluting in 25 µl RNase-free water. To determine the optimal cycle number for library amplification, the amplification curve inflection point was determined for each sample by qPCR. 2.5 µl of each sample was added to a master mix of 1 µl 2.5X Thiazole Green (Biotium, 40086), 1 µl 2.5X ROX reference dye (Thermo Fisher Scientific, 12223-012), 12.5 µl Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, M0494S), 1.25 µl 10 µM IDT_i7_1 primer for plasmid amplicons/P7_universal_reverse for viral single-stranded cDNA, 1.25 µl 10 µM IDT_i5_1 primer and 5.5 µl of molecular grade water. The following thermocycler conditions were used: 98 °C - 3 mins, followed by 40 cycles of 98 °C - 10 s, 68 °C - 20 s, 72 °C - 30 s.

Subsequently, the remaining 22.5 µl of each sample was added to a mix of 25 µl Q5 Hot Start High-Fidelity 2X Master Mix, 2.5 µl 10 µM IDT_i7_1-10 primer for plasmid amplicons/P7_universal_reverse for viral single-stranded cDNA, and 2.5 µl 10 µM IDT_i5_1-10 primer. The IDT_i5_X and IDT_i7_X primers each contain 8 bp index sequences derived from the rhAmpSeq index primer set (Integrated DNA Technologies) for dual-index demultiplexing alongside the Illumina P5 and P7 adapter sequences, respectively. Thermocycling conditions for this Illumina adaptor PCR were: 98 °C - 3 mins followed by X* cycles of 98 °C - 10 s, 68 °C - 20 s, 72 °C - 30 s, then 72 °C - 2 mins. *Where X is the number of cycles to reach the qPCR inflexion point, minus 3 cycles to account for the 10X (2^3) increase in target concentration.

For viral amplicons, the P7-indexed adaptor sequences were added in a final PCR step. The amplicons were first gel purified with a Wizard SV Gel and PCR Clean-Up kit (Promega, A9281) and eluted with 20 µl water. Then, the same PCR mix as the Illumina adaptor PCR was set up, substituting the P7_universal_reverse primer for the IDT_i7_X primer matching the IDT_i5_X for each sample. Thermocyling conditions were likewise the same as the Illumina adaptor PCR but with a fixed number of 3 cycles.

The indexed viral/plasmid amplicons were then gel purified with a Wizard SV Gel and PCR Clean-Up kit (Promega, A9281) and eluted with 25 µl water. The concentration and size distribution of these amplicon libraries were quantified with a High Sensitivity D1000 ScreenTape kit (Aligent, 5067-5584) on an Aligent 2200 TapeStation system. Samples passing QC were sequenced on an Illumina NovaSeq 6000/NovaSeq X using Illumina Trueseq primers for paired-end 100 bp sequencing at a depth of 10-50 million reads per sample.

2.3.4 Quantification of barcode library diversity

To analyse sequenced libraries, low-quality reads were first removed from raw demultiplexed .fastq files with Trimmomatic 0.36 (removing reads with an average PHRED score per base of less than 20 in a 6 base sliding window and removing any reads with a length of less than 101 bp). The sequences flanking the barcode site were then matched to their expected sequences, and any mismatching sequences were discarded to remove frameshifts. Next, the UMI sequences were used to de-duplicate the remaining sequence reads to remove PCR amplification bias. The remaining sequences represented individual species of barcode RNA/DNA found in the viral/plasmid samples. To account for sequencing errors and PCR-induced mutations that may be present in some of these sequences, closely related barcodes were collapsed into their most abundant "parent" sequences. This process of barcode collapse has been modified from a similar analysis performed on Sindbis viral libraries in the MAPseq barcode library analysis pipeline (Kebschull, Garcia da Silva, Reid, et al., 2016). First, Bowtie2 (2.5.1-GCC-12.3.0) was used to align viral barcodes to all other barcodes in the library with a maximum hamming distance of 2. The UMI counts of the barcodes within each "family" of related sequences were then used to collapse them into the most abundant barcode, summing together all of the counts.

2.4 Results

2.4.1 Library production

Several viral libraries will be described in the following section. I performed barcoding of the rabies genome plasmid using a modified version of an inverse PCR / ligation protocol (Methods 2.3.1) previously described for barcoding rabies viruses (Saunders et al., 2021). The glycoprotein-deleted CVS-N2c genome plasmid was modified to include the full-length hepatitis delta virus antigenomic ribozyme (HDVagrz) for improved viral rescue efficiency (Ghanem et al., 2012, Methods 2.3.1, Figure 2.1 A). Barcodes comprised of 20 random nucleotide bases were then incorporated downstream of the mCherry transgene region of CVS-N2c strain rabies plasmids using oligonucleotide primers (Figure 2.1 B, Methods 2.3.1). The resultant linear PCR products were restriction digested (Figure 2.1 C) and their sticky ends were ligated together to form open circular barcoded plasmids (Figure 2.1 D).

These pools of open circular plasmids were transformed into electrocompetent *E. coli* and plated onto large bioassay plates to form lawns of colonies (Figure 2.2 A). The amplified genome plasmids extracted from these *E. coli* lawns were transfected into HEK293T cells along with plasmids expressing CVS N2c N, P, L and G genes as well as T7 polymerase (Figure 2.2 B). After initial viral rescue and amplification, viral supernatant was trans-

ferred to BHK-EnvA cells for pseudotyping (Figure 2.2 C). Barcode sequences from the plasmid and from glycoprotein-coated and pseudotyped rabies viral stages were isolated and a series of nested PCR steps (Methods 2.3.3) were used to produce amplicon libraries for Illumina sequencing to determine the barcode abundance distribution.

During viral library production, it is possible for two forms of rabies virus to be present in the final supernatant. I compared the measured titres of these two viral populations for each viral library that was produced (Figure 2.3). One of these components was the desired population of EnvA-pseudotyped viral particles. These pseudotyped viruses are only capable of infecting cells expressing the TVA receptor protein, which is not present in mammalian cells thus making them targetable towards a defined population of starter cells during a barcode labelling experiment. The other possible component in the final viral libraries was contaminant glycoprotein-coated viral particles, left over from the initial supernatant infection of the BHK-EnvA cells. These viruses would be capable of initiating infection of random non-starter cells near the injection site. If the titre of these glycoprotein-coated viral contaminants were too high then this risked complicating connectomic analyses using these libraries. A discussion of the problems that these contaminant infection events pose is described following from Figure 5.13, where control

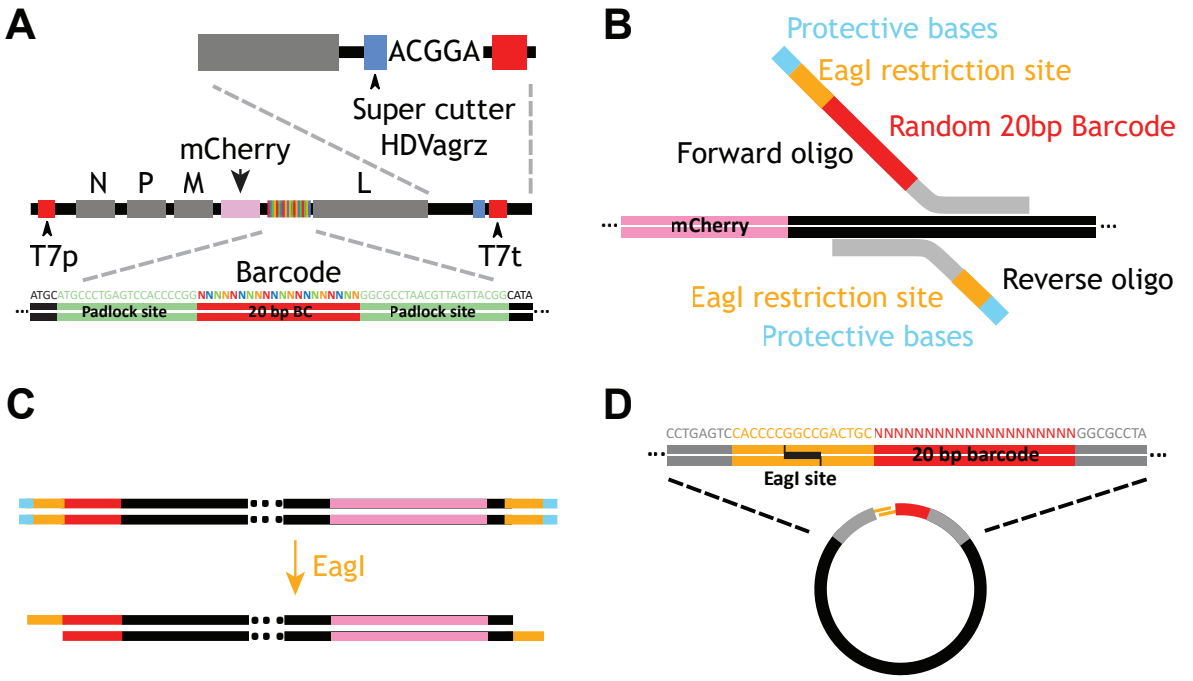


Figure 2.1: Generating barcoded rabies genomic plasmids. **(A)** Diagram of the barcoded rabies genome with barcodes inserted downstream of the mCherry transgene region, flanked by padlock binding sites (green). At the 3' end of the genome, the extended sequence of the super cutter hepatitis delta virus anti-genomic ribozyme (HDVagrz) has been inserted. **(B)** Barcode insertion site downstream of the mCherry transgene, with inverse PCR primers for barcode incorporation. **(C)** Barcoded linear PCR products with terminal EagI restriction sites (yellow) and protective terminal bases (blue) are digested to produce sticky ends. **(D)** PCR products are ligated together to form open circular barcoded plasmids.

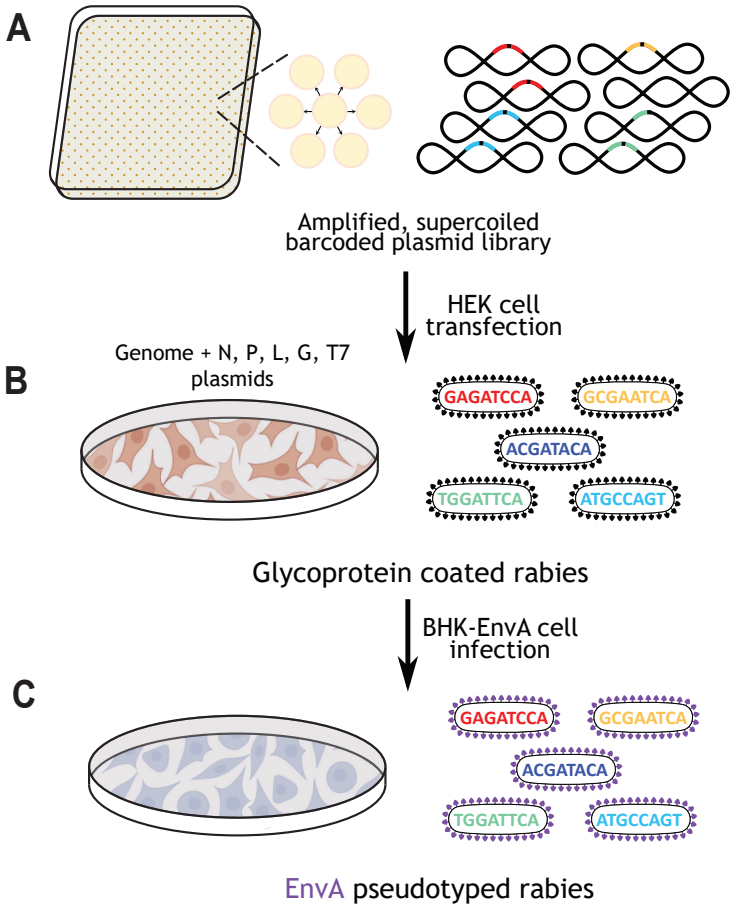


Figure 2.2: Rabies virus plasmid transformation, transfection and rescue. (A) Open circular barcoded plasmids are transformed into *E. coli* where they are amplified and supercoiled in plate culture conditions, colony growth is restricted by neighbouring colonies limiting skewed clonal expansion. **(B)** Supercoiled genome plasmids are transfected into HEK293T cells with rabies gene plasmids N,P,L,G and a T7 polymerase plasmid to rescue barcoded viruses. **(C)** Supernatant from HEK293T containing G-coated rabies is transferred to BHK-EnvA cells for EnvA pseudotyping.

infection experiments were carried out in mouse brains to determine the prevalence of these contaminant infection events in the same conditions as *in vivo* connectomic experiments. For all libraries pseudotyped using the protocol described here, the ratio of contaminant to pseudotyped viral particles was over 1:10,000 (Figure 2.3).

2.4.2 Quantification of barcode library diversity

Quantification of the barcode libraries extracted from plasmid DNA or viral RNA was performed by counting the abundance of barcode sequences within a sample using Illumina sequencing. Extracted barcode sequences were tagged with 12 bp unique molecular identifiers (UMIs) during PCR-based library preparation. As the quantification of barcode sequence abundance requires an accurate measure of molecule counts, the incorporated UMI sequences were necessary to mitigate the effects of PCR amplification bias on these libraries. After quality control filtering of Illumina reads and removal of misprimed reads

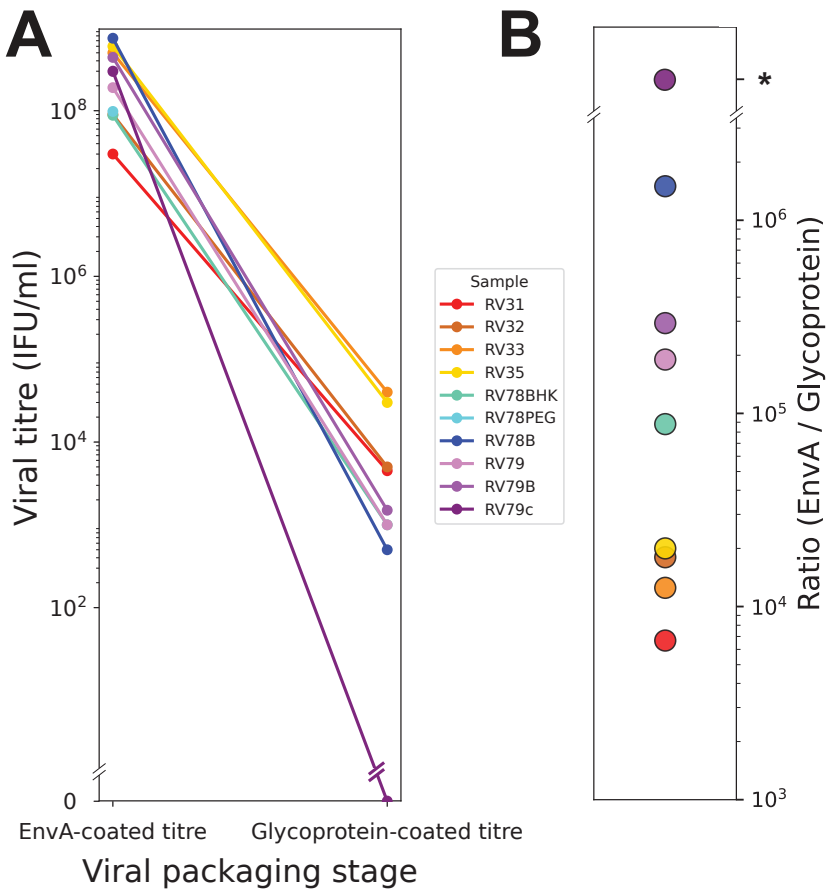


Figure 2.3: Viral library titres. **A** Titre of EnvA-pseudotyped viruses and glycoprotein-coated viruses in the final viral library concentrates after ultracentrifugation. Titres were determined by fluorescent cell counting after serial dilution onto HEK293T-TVA and HEK293T cells respectively. **B** Ratio of EnvA coated to glycoprotein-coated viruses (asterisk denotes samples with no detected glycoprotein-coated infection).

by matching the positions of known flanking sequences surrounding the barcodes (Methods 2.3.4), identical barcodes also sharing the same UMI were mapped to a single input molecule and their counts were deduplicated (Figure 2.4). The mean sequencing depth across samples (pattern-matched reads / de-duplicated reads) was 9.0±8.9 (mean±standard deviation). However, RV31, RV35 and plasmid PBC20 were all under-sequenced compared to the other samples, with sequencing depths of 1.9, 2.1 and 1.9, respectively. This means that the diversity measured from these samples is an underestimate of the total diversity of sequences. This does not affect the calculation of the expected fraction of uniquely labelled cells from these libraries, as this is based on the barcode abundance distribution rather than the absolute number of reads per barcode. Additionally, to eliminate misinterpretation of barcode sequences that were mutated from their original sequence during library PCR amplification or had bases miscalled during Illumina sequencing, "barcode collapse" was performed on families of neighbouring sequences. Barcode sequences within a certain Hamming distance of each other were grouped into families, and their molecule counts were merged into that of the most abundant parent barcode to "collapse" errors into this sequence that they were likely to originate from (Figure 2.4).

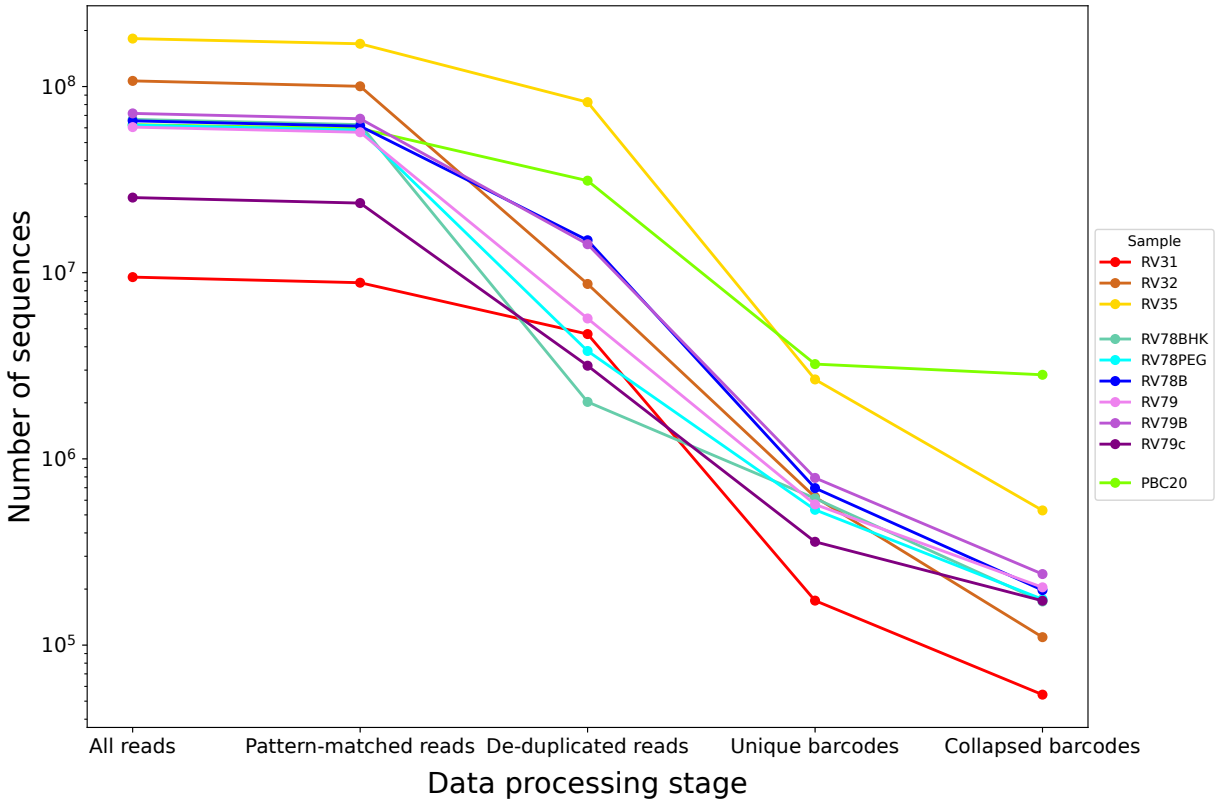


Figure 2.4: De-duplication and collapse of library sequences. Sequence counts for each Illumina sequenced barcode library at each stage of data processing. Counts from All reads show the total read counts from each sequencing run before applying any quality control filters. Counts from Pattern-matched reads show the total read counts after performing Trimmomatic filtering based on PHRED scores, followed by filtering out bad sequence reads that don't match known sequences flanking the barcode region (Methods 2.3.4). De-duplicated reads show the total sequence counts after removing whole sequence duplicates from PCR amplification. Unique barcodes show the total number of unique barcodes after de-duplicating reads with the same UMI sequence. Collapsed reads show the total number of unique barcodes after performing edit-distance collapse to remove PCR/sequencing errors with a hamming distance of 2 (Methods 2.3.4).

As the number of unique sequences found within these barcode libraries ranged from only 10^5 - 10^6 sequences in a possible 20 bp barcode space of 4^{20} (1.09 x 10^{12}) sequences, the likelihood of barcodes collapsing at random into neighbouring sequences should be low. This rate of random collapse was estimated by generating a random set of barcode sequences equal in number to the number of unique barcodes in the RV35 barcode library, and then calculating the distribution of their minimum Hamming distances to one another. This was compared with the minimum Hamming distance distribution of the unique barcodes in the RV35 barcode library, before and after error correction collapse (Figure 2.5). For a set of random 20 bp barcodes, only 0.08% are estimated to have a minimum Hamming distance of two or fewer to another barcode. Before collapse, 91.6% of the unique RV35 sequences had a minimum Hamming distance of two or fewer to another barcode, indicating an abundance of single base errors which were then corrected to their high abundance "parent" sequences. These errors can be attributed to PCR library amplification and sequencing errors as well as the high mutation rate of RNA virus

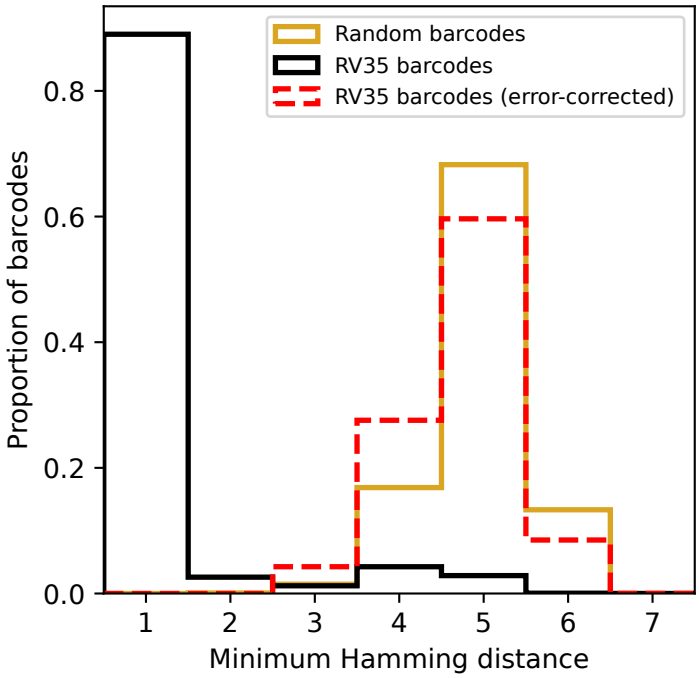


Figure 2.5: Error correction of barcode libraries. The distribution of minimum Hamming distances for all unique barcodes in the RV35 barcode library before (black), and after (red) error correction with a Hamming distance of 2, compared with a reference distribution of the minimum Hamming distances for a random barcode set of equal size to the RV35 library (yellow).

polymerases compared to DNA polymerases (Combe et al., 2015; Hanada et al., 2004; Sanjuán et al., 2010).

After performing these de-duplication and barcode collapse steps, the abundance distributions of the barcode libraries were compared. Due to the low rescue efficiency of rabies viruses, there was a drop in the total number of unique barcode sequences observed from the plasmid barcode library to the viral barcode libraries. The barcoded plasmid library that was selected for the packaging of all subsequent viral libraries (library PBC20) contained a total of 2,867,860 unique barcodes after error correction. The total abundance of the 100 most abundant barcodes from the plasmid library constitutes 0.06% of all barcode reads from the library. Comparing the plasmid library to the most diverse viral library that was produced from all packaging experiments (RV35), the number of unique barcodes observed was reduced over five-fold (Figure 2.6 A, PBC20 2.8 x 10^6, RV35 5.3 x 10^5) whilst the total abundance of the 100 most common barcodes in the library rose 65 fold to 3.9% (Figure 2.6 B). Using the abundance distribution of the barcodes within each library, the probability of selecting each barcode within the library was calculated and used to estimate the percentage of cells which would be uniquely labelled for any given number of independent infection events using that library (Equation 2.2.2). As expected, the decrease in total unique barcodes and the increased skew of barcode abundance distribution from the plasmid library to the viral library resulted in a decrease in the percentage of cells that are predicted to be uniquely labelled (Figure 2.6 C). If the plasmid library were to be used to label 2000 cells directly, the expected percentage of

uniquely labelled cells (Equation 2.2.2) would be 99.9%. The RV35 barcode library is still highly diverse, and would be expected to uniquely label 93.0% of 2000 cells.

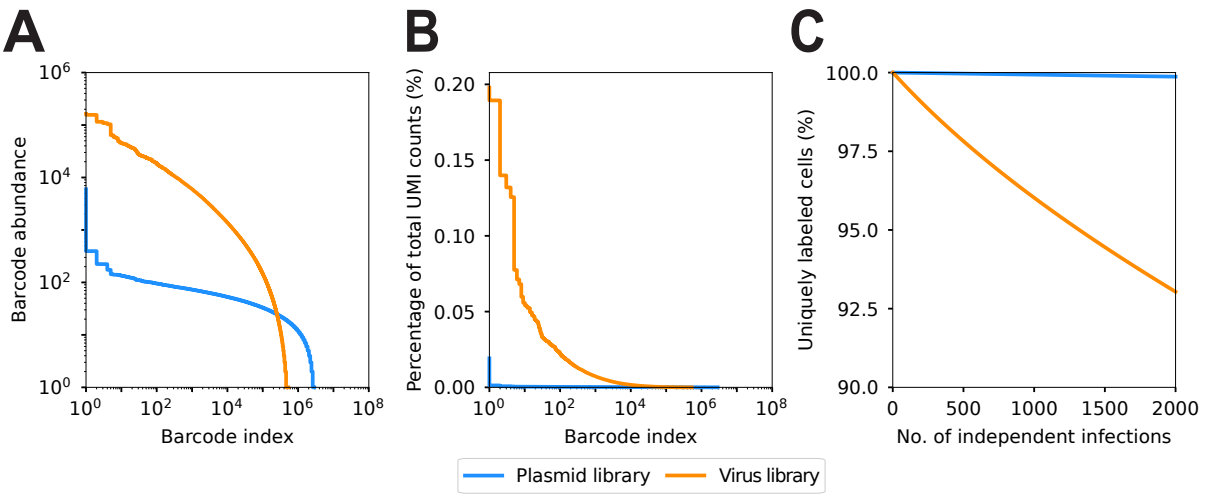


Figure 2.6: Library diversity loss during viral packaging. **(A)** Distribution of UMI counts per unique barcode sequence found in the barcoded plasmid library (blue) and in the final viral preparation of EnvA pseudotyped barcoded virus (orange). **(B)** As in (A) but barcode UMI counts have been normalised as a percentage of the total UMI counts in the respective libraries. **(C)** Calculated fraction of infected cells which would be labelled with a unique barcode sequence, given the distribution of barcode abundances in the barcoded plasmid library PBC20 (blue) and in the final viral preparation of EnvA pseudotyped barcoded virus RV35 (orange).

To determine the contribution of the pseudotyping step in the loss of barcode library diversity, supernatant samples were taken during the packaging of a viral library the day before the pseudotyping stage (RV78BHK), and then again after the pseudotyping stage at day 3 (RV78PEG), 5 (RV79B) and 6 (RV79C) post transfer. The viral barcode sequences from these time points were isolated and sequenced to determine the changes in barcode abundance distribution (Figure 2.7 A, B). The total abundance of the 100 most common sequences increased by 42.3% between the glycoprotein-coated viral time point and the first pseudotyped viral time point, but only by 1.3% from the first to last pseudotyped time points. Likewise, the estimated fraction of uniquely labelled cells decreased from 91.6% for 2000 cells at the glycoprotein-coated viral time point to 87.6% at the first pseudotyped viral time point, but then remained stable at 87.6%, 87.4% for the later pseudotyped time points (Figure 2.7 C).

The effect of the scale of initial viral rescue (total surface area of cell culture) on library diversity was also explored by sampling the final pseudotyped viral libraries of two viral rescues initiated from transfection of two 6-well plate wells each (RV31 and RV32), and a viral rescue initiated from transfection of twelve 6-well plate wells (RV35). The difference in barcode abundance distribution between the 2 well and 12 well conditions was large. The total number of unique barcode sequences increased 6.4-fold from a mean of 82,228 for the 2 well condition, to 528,774 for 12 wells (Figure 2.8 A). The total abundance of the 100 most common sequences decreased 3.7-fold from 14.6% to 3.9% (Figure 2.8 B).

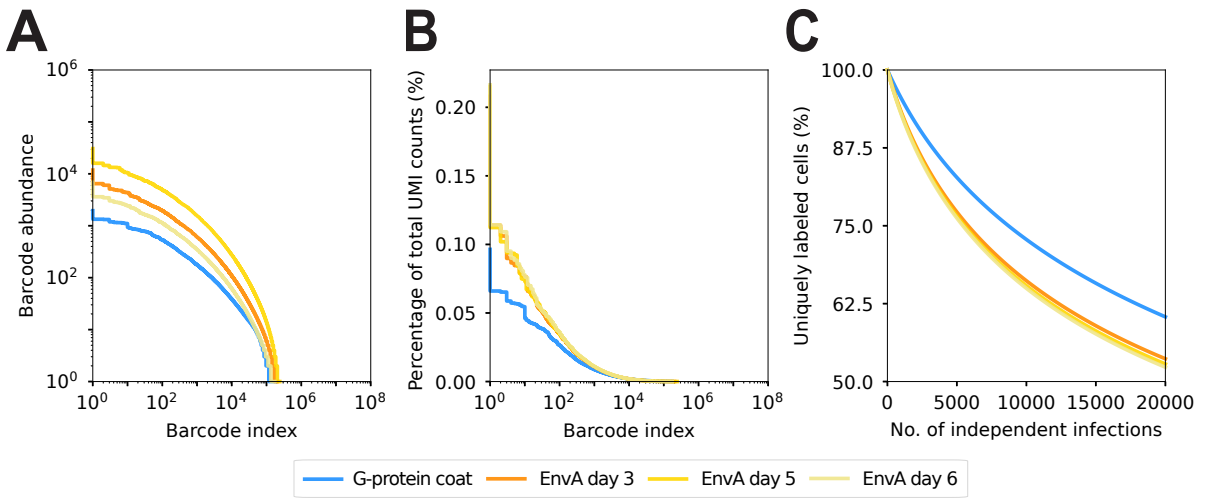


Figure 2.7: Library diversity loss during pseudotyping. (A) Distribution of UMI counts per unique barcode sequence found in a glycoprotein-coated viral library (blue) and in the EnvA pseudotyped barcoded viral libraries collected at various time points (shades of orange), using the same glycoprotein-coated library to begin pseudotyping. **(B)** As in (A) but barcode UMI counts have been normalised as a percentage of the total UMI counts in the respective libraries. **(C)** Calculated fraction of infected cells which would be labelled with a unique barcode sequence, given the distribution of barcode abundances in the glycoprotein-coated viral library (blue) and in the EnvA pseudotyped barcoded viral libraries (shades of orange).

This increased the estimated fraction of uniquely labelled cell from a mean of 66.8% for 2000 cells for the two well condition to 93.0% for the 12 well condition.

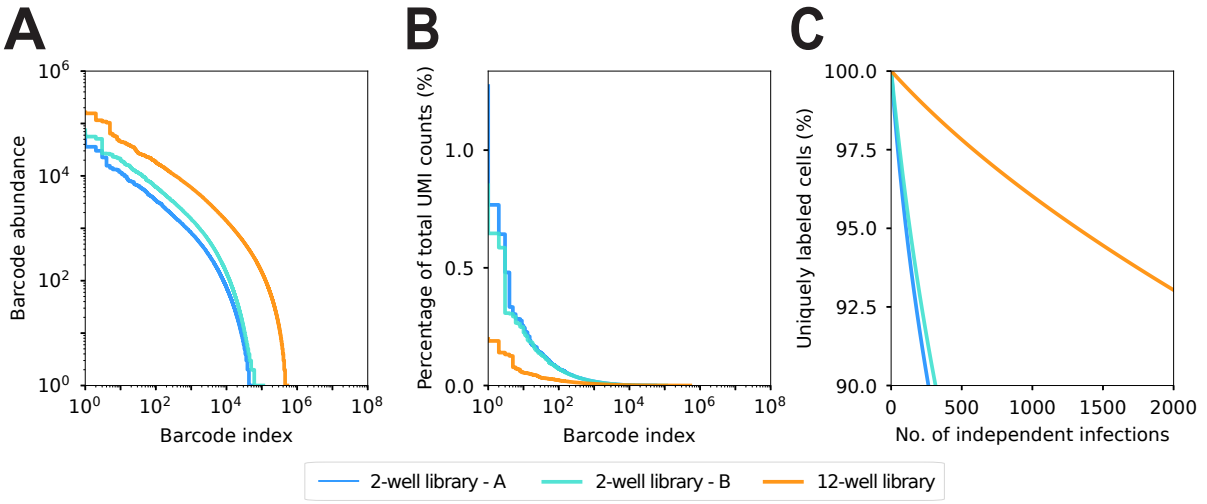


Figure 2.8: Library diversity scaling with rescue size. (A) Distribution of UMI counts per unique barcode sequence found in two replicate EnvA pseudotyped barcoded viral libraries which started rescue from two 6-well plates (shades of blue) and an EnvA pseudotyped barcoded viral library which started rescue from twelve 6-well plates. **(B)** As in (A) but barcode UMI counts have been normalised as a percentage of the total UMI counts in the respective libraries. **(C)** Calculated fraction of infected cells which would be labelled with a unique barcode sequence, given the distribution of barcode abundances in the replicate two 6-well plate viral libraries (shades of blue) and in the twelve 6-well plates viral library (orange).

2.5 Discussion:

In comparison to the relative ease of producing barcoded viral libraries from RNA retroviruses (Fennell et al., 2022), positive single-stranded RNA viruses (Kebschull, Garcia da Silva, and Zador, 2016) and single-stranded DNA viruses (Davidsson et al., 2016), producing barcoded libraries from negative-strand rabies viruses is challenging. By optimising established library production methods to achieve an efficient means of barcode incorporation, transformation and rescue, I aimed to minimise barcode loss during library production. Likewise, by restricting bacterial growth with plate culture and reducing viral passaging during packaging stages, I aimed to minimise the extent of biased clonal expansion. To ensure that these precautions were effective in producing viral libraries of sufficient diversity for uniquely labelling most starter neurons in the conditions expected of an *in vivo* infection event, it was necessary to quantify the diversity of these libraries. As has been evidenced in the past, the inefficiency of rabies viral rescue from plasmid DNA inevitably leads to loss of library diversity from the barcoded plasmid to viral stages (Clark et al., 2021; Saunders et al., 2021). The same is true of the highly diverse plasmid library which I have generated here, and the comparatively skewed abundance distributions of the viral libraries that have been generated from it. However, when compared to the libraries generated from other recent rabies barcoding efforts, the diversity of the rabies libraries described in this thesis exceeds their ability to uniquely label cells and demonstrate one of the highest diversity rabies barcode libraries that have been produced to date. Whilst the analysis performed on the SBARRO barcoded rabies libraries (Saunders et al., 2021) was extensive and influenced some of the analyses performed on my own libraries, other libraries were insufficiently analysed to allow for true quantification of barcode diversity and expected unique labelling capability. The libraries described in the RABID-seq barcoded rabies approach (Clark et al., 2021) contain a semi-random 28 bp barcode sequence flanked by PCR handles for specific PCR enrichment during scRNAseq library preparation. Rather than attempting to minimise clonal amplification in the rabies barcode packaging stage, standard rabies packaging protocols were used to produce these barcoded viral libraries. The abundance distribution of the resultant barcoded libraries was also not taken into account when performing estimations of the percentage of uniquely labelled cells which were achievable with the libraries. Instead, the estimated uniquely labelled cells were calculated based solely on the number of unique barcode sequences detected in the library. Likewise, no abundance distribution analysis was performed in a more recent study that used BARseq to detect barcoded rabies viruses (Zhang et al., 2024). This points towards the need for a formalised set of analyses to be performed on future barcoded rabies viral libraries, given the high importance of barcode diversity in producing unambiguous connectomic information.

Given that the ideal conditions of the *in vivo* experiments described in later chapters of this thesis (see Chapter 5.4.1 for starter cell conditions) would result in labelling from

hundreds to around two thousand starter cells in a single experiment in the visual cortex, the highest diversity pseudotyped viral library described here (RV35) is already capable of labelling over 92% of a population of 2000 starter cells. For future experiments that call for the ability to label larger populations than this, the increase in library diversity that was seen from scaling up the number of transfected cells used in the rabies rescue, promises to allow for the generation of improved libraries in the future by following the protocol described here. Further improvements may be possible by utilising stable transfectable cell lines that express more of the proteins that are necessary for initial rabies rescue events (Shin et al., 2024). Increasing the efficiency of rescue may also be possible by further modifications to the rabies rescue plasmid, incorporating modifications to the Hammerhead ribozyme sequence which has demonstrated a 3-fold improvement in rabies rescue (Mori and Morimoto, 2014; Morimoto et al., 1998).

Chapter 3

Scaling up in situ sequencing

3.1 Authorship declaration

I, Alexander Becalick, performed all of the experiments and analyses in this chapter. Xavier Cano-Ferrer and George Constantinou (members of the Making Lab at the Francis Crick Institute) built the control board and resistive-heating units for the in-line pre-heater and flowcell heater of the microfluidic setup.

3.2 Introduction

In order to capture the connectivity of the neurons in a barcode labelling experiment, it is necessary to detect the sequences of these barcodes and attribute them correctly to their source cell. Determining a complete connectivity matrix for the labelled cells requires all of them to be detected. Here I describe a method to extract this information from large volumes of tissue in a time-efficient manner.

Single-cell RNA sequencing is one way to read barcodes from individual rabies-infected cells (Clark et al., 2021; Saunders et al., 2021; Zhang et al., 2024). This method can utilise the speed of high throughput sequencing platforms to provide a readout of cell-barcode information. However, a large proportion of neurons are lost during the tissue dissociation and cell sorting processes that are necessary for single-cell sequencing approaches (Saunders et al., 2021). This means that the resultant data is missing both presynaptic and postsynaptic neurons. As well as reducing the effective throughput of the experiment by losing a large amount of connectivity data, this cell loss can also result in ambiguity in the connectomic information. This is due to the possibility that multiple "starter" cells that initiate monosynaptic rabies spread to presynaptic neurons, can share an identical barcode (see Chapter 5.4.1 and 5.5 for a description of the scenarios that can result in this). The connectivity of the presynaptic cells with this shared starter cell barcode cannot be unambiguously assigned to either of the starter cells. If large proportions of all

cells are lost during tissue dissociation and cell sorting, these identically barcoded starter cells cannot be reliably detected and removed from connectomic analysis.

An alternative to single-cell RNA sequencing is to perform sequencing in situ, directly on tissue from the viral injection site. Dissociation of the tissue into individual cells is not required for in situ sequencing, so cell loss can be mitigated if the whole region of interest is instead sampled by slicing into sections for sequencing. The landscape of spatial transcriptomic techniques has expanded rapidly in the past decade from multiplexed extensions of fluorescent in situ hybridisation (FISH) techniques (Chen et al., 2015; Eng et al., 2019; Jakt et al., 2013), including various signal amplification approaches like branched DNA amplification (Battich et al., 2013; Wang et al., 2012) and hybridisation-chain reaction (Choi et al., 2014; Dirks and Pierce, 2004; Shah et al., 2016), to spatially barcoded RNA array-capture techniques which rely upon traditional high throughput sequencing platforms to sequence transcripts with added spatial information (Chen et al., 2021; Kim et al., 2024; Rodriques et al., 2019; Vickovic et al., 2019).

There are several key requirements of a barcoded connectomic experiment. These are as follows: the technique must allow for the readout of arbitrary random barcode sequences, the technique must be capable of assigning transcripts to individual cells, and the technique must be scalable to large amounts of tissue. Readout of arbitrary sequences is not achievable by techniques such as seqFISH (Shah et al., 2016), MERFISH (Chen et al., 2015), coppaFISH (Bugeon et al., 2022), CosMx (He et al., 2022) and other techniques that use readout probes designed to detect pre-determined barcode sequences. This arbitrary readout should, in practice, be achievable by existing array-based sequencing techniques such as Slide-Seq (Rodriques et al., 2019), HDST (Vickovic et al., 2019), Stereo-Seq (Chen, Liao, et al., 2022), Seq-scope (Kim et al., 2024) and DBiT-Seq (Liu et al., 2020). However, each of these techniques lacks the spatial resolution that is required to provide true single-cell barcode readout. This is due to limitations of the size of spatial units that the arrays are composed of in the case of Slide-Seq (10 µm spheres) and DBiT-Seq (10 µm channels), or due to lateral diffusion of captured RNA transcripts in Visium, Stereo-Seq and Seq-scope.

Achieving subcellular resolution is necessary for assigning barcodes accurately to individual cells in densely crowded tissue. The ability to sequence arbitrary barcodes whilst resolving the subcellular location of individual transcripts is achievable using the imaging-based in situ sequencing technique - BARseq (Sun et al., 2021); which uses oligonucleotide padlock probes which bind to the cDNA of reverse-transcribed transcripts. These padlocks are designed to bind to regions flanking either side of the viral barcode cDNA so that a polymerase can be used to copy the barcode sequence from the cDNA into the gap between the padlock arms, before ligating the padlock into a circular piece of DNA (see Chapter 5, Figure 5.7). This padlock can be amplified by the highly processive phi29 polymerase to create a ball of DNA (termed a rolony) containing many copies of the rabies

barcode sequence. Rolonies have an average diameter of 800 nm as determined by electron microscopy (Deng et al., 2018) and can be detected by performing successive rounds of sequencing-by-synthesis - a polymerase induced incorporation of fluorophore-conjugated nucleotides to sequence a region of interest (Ju et al., 2006). By using 4-colour Illumina sequencing reagent-derived nucleotides, these barcode sequences can be read out with fluorescence microscopy to resolve transcripts at a near-diffraction-limited spatial resolution (Sun et al., 2021). The nucleotides in an Illumina sequencing kit contain an azido terminator moiety to prevent the addition of more than a single base per sequencing round, along with a conjugated fluorophore which is different for each nucleotide: A, T, G and C. Both of these additional molecular groups are attached to the nucleotides with chemically cleavable linkers. Sequencing is performed in cycles of nucleotide incorporation, imaging and then dye/terminator cleavage of these nucleotides to allow for the next nucleotide to be added to the growing strand of sequenced DNA.

BARseq is typically performed on tissue sections up to 20 µm thick, and the helper AAV viruses that are used to define the initially infected "starter" cells in a rabies transsynaptic experiment spread approximately 500 µm radially from the site that they are injected. This means that an in situ barcoded connectomics experiment would require sequencing at least 50 tissue sections to cover the approximately 1 mm thick block of tissue that has the potential to contain infected starter cells. As each of these tissue sections must undergo many rounds of imaging and sequencing chemistry to complete a sequencing-by-synthesis experiment, I had to achieve the highest sequencing throughput possible to minimise the time taken to complete each experiment. I did this by building a sequencing and imaging setup capable of imaging at high speeds and with the goal of completing intermediate chemistry steps between each imaging round in an automated manner to reduce hands-on time.

3.3 Results

3.3.1 Development of a beamsplitter for simultaneous multi-channel imaging

In a single-camera microscope setup, imaging all four nucleotide dyes requires multiple image acquisitions with filter changes for each dye's excitation/emission wavelength, limiting imaging throughput. A four-way beamsplitter was necessary to eliminate this filter movement and allow all four nucleotide dyes to be imaged at once. I achieved this by placing a main dichroic mirror within the excitation/sample path and using a set of three emission dichroic mirrors to split the emission light path towards four separate cameras (Figure 3.1 A). The beamsplitter design consists of the following elements: first, a tube lens was used to produce an infinity-corrected optical region in which to place the dichroic

mirrors and emission filters to reduce aberrations from the interaction of non-collimated light (Greb, 2016; Sarkar, 2022). This was paired with identical tube lenses placed before each camera to form an image. Within the infinity-corrected space, three adjustable dichroic mirror mounts were placed to enable fine alignment of the four images onto the cameras. Dichroic mirror holders were machined to hold the mirrors without causing curvature. Emission filter holders were placed in each of the four camera light paths and a combination of rotatable camera mounts and length-adjustable tubing enabled the scale and rotation of each camera image to be matched (Figure 3.1 B). A full parts list is available in Appendix 7.1.

To optimise the light-collecting capability of the beamsplitter, the transmission wavelengths of the dichroic mirrors and emission filters were matched to the nucleotide dyes. This was confounded by the closed-source nature of the Illumina sequencing reagents. No documentation of the emission spectra for the nucleotide-conjugated dyes from the sequencing kits was publicly available, only the excitation wavelengths used by the Illumina sequencers. To determine these emission spectra, I used a microscope with a tunable prism-based spectrometer (Leica SP8, Leica microsystems) to perform a lambda scan through 3 nm emission wavelength bins to find the peak emission wavelength of each dye (Figure 3.2). I then selected dichroic mirrors and emission filters which would capture as much of each dye's emission as possible whilst minimising bleedthrough between the four channels.

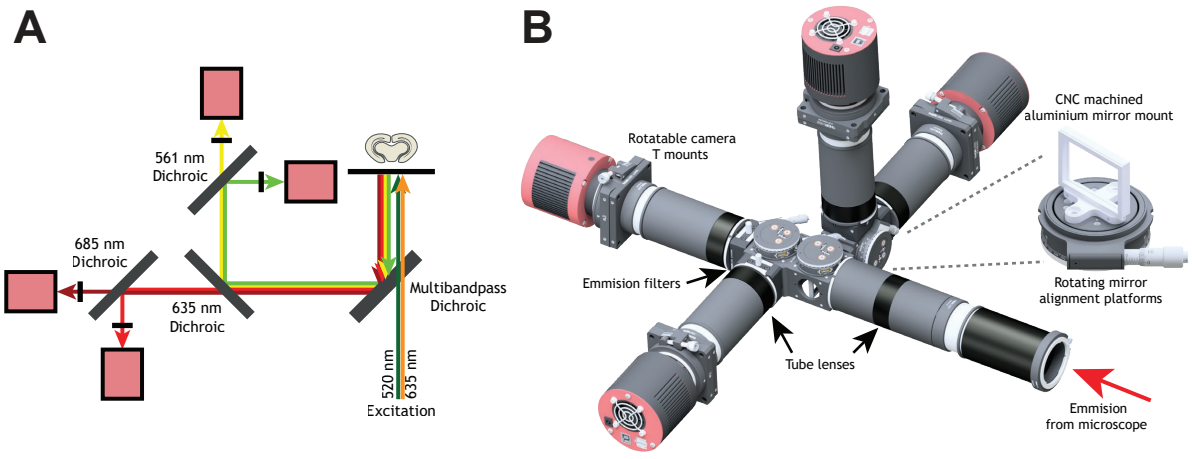


Figure 3.1: 4-way beamsplitter layout for simultaneous imaging. **(A)** A diagram of the light path of the in situ sequencing microscope. A multi-bandpass dichroic within the microscope body allows for sample excitation with multiple wavelengths. Emission light is split between the four cameras with three emission dichroics. Each camera also has an emission filter placed after the dichroics. **(B)** A 3D-CAD design of the custom beamsplitter. Composed of Thorlabs components and CNC-machined mirror mounts, the system has XY mirror tilt, mirror rotation and camera rotation adjustment.

Microscope setup

The beamsplitter was paired with a Nikon Ti2-e inverted microscope body, equipped with a PZ-2000 XYZ piezo stage (Applied Scientific Instrumentation). Fast Z-stack acquisition was achieved using TTL triggering from the exposure out signal of one of four CellCam Kikker 100MT cameras (Cairn Research). Two microscope systems were built, one using a commercial 4-way MultiCam beamsplitter (Cairn Research) and the other using the custom beamsplitter assembled from Thorlabs components, as described in Figure 3.1. All other parts were the same between the two imaging systems. All tissue imaging was performed using a 20x 0.8 NA Lambda-D air objective, with a CoolLED pe-4000 light source. Images were captured with 0.231 µm pixel size in 23 plane Z-stacks with a Z spacing of 1 µm and tile overlap of 10%. All microscope hardware was controlled using micromanager as a GUI interface (Edelstein et al., 2010), with pycromanager (Pinkard et al., 2021) being used to automate the acquisition of multiple flowcells per round.

Beamsplitter imaging throughput

To compare the difference in acquisition time between using a four-camera beamsplitter and using a single-camera system, the microscope was set to acquire images in both configurations and imaging metadata was used to calculate the mean time to acquire a single tile position, including stage movement. In the single camera mode, the microscope would acquire a full Z-stack for one channel, then acquire the other three channel Z-stacks. The microscope was later upgraded to enable hardware triggering from the cameras, resulting

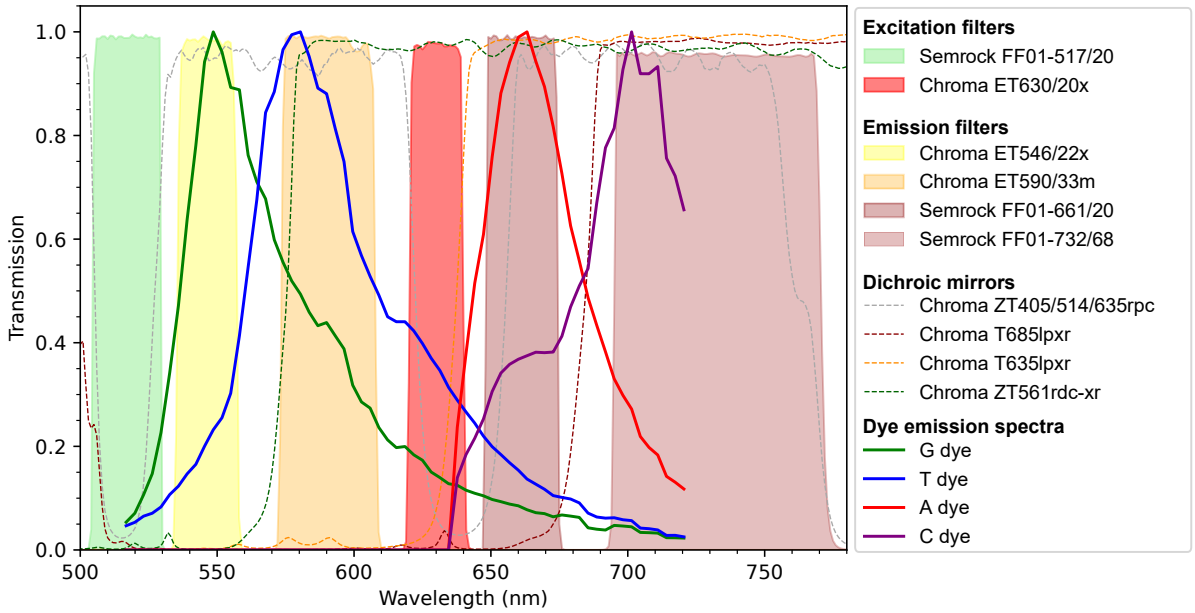


Figure 3.2: Dye emission spectra and optimised filters. Emission spectra of each nucleotide dye normalised to their peak emission wavelength (green, blue, red, and magenta lines) overlaid onto the transmission values of the selected dichroic, excitation and emission filters used for the microscope beamsplitter.

in the movement of the piezo-Z stage to the next Z-stack position once the cameras had finished exposing the current frame. Comparisons using software stage triggering and hardware triggering were made in the beamsplitter mode. A single-tile acquisition in the single-camera mode was measured to take 37.2 seconds (Figure 3.3 A). This would result in a 20.6-hour acquisition time for a single sequencing round of 2000 tile positions, which is the amount of tiles expected from imaging four flowcells (see 3.3.3), each with 10 ROIs that cover a third of a mouse coronal brain section per ROI. In the comparable beamsplitter mode without hardware triggering, a single tile acquisition took 10.8 seconds, a 3.44X improvement (6.0 hours to acquire 2000 tiles). These estimates do not take into account the added time that a single-camera system would require for filter switching between channel acquisitions, and so represents a lower limit on the real-world performance gain of a beamsplitter. Perfect scaling to four times the imaging speed of a single camera setup wasn't achieved due to a fixed time delay present between tile acquisitions from the stage XY movement and an autofocusing routine that occurred at each tile position. Hardware-triggering of the Z-stack movement removed software delays between the camera exposure completion and stage movement, decreasing the single tile acquisition time further to 5.6 seconds (3.1 hours to acquire 2000 tiles). This was a 1.92X improvement over software triggering. The theoretical limit of the imaging speed of the current hardware-triggered beamsplitter setup, disregarding stage movements, autofocusing and software delays is 3.36 seconds per tile (21 Z-positions with a 160 ms exposure time). This time could also be decreased further by using a brighter light source to decrease the exposure time.

To produce an estimate of the total time required for each imaging configuration to complete a full sequencing experiment with seven rounds of gene sequencing and ten rounds of rabies barcode sequencing, I used the time that it takes to complete a single round of 2000-tile imaging and added a fixed 45 minutes per round for the time that it takes to complete sequencing chemistry between each imaging step. In single-camera mode, this results in a 15.2-day sequencing experiment (Figure 3.3 B). This is reduced to a 4.8-day experiment with a software-triggered beamsplitter and reduced again to a 2.8-day length experiment with a hardware-triggered beamsplitter.

3.3.2 Beamsplitter performance comparison to a commercially available system

Whilst using a beamsplitter does vastly improve the speed of the imaging setup, the introduction of a pair of tube lenses, three dichroic mirrors, emission filters and the reduced diameter apertures between the mounts of these components may introduce aberrations and vignetting to the system. To assess the extent of these effects, the image plane of the Thorlabs custom beamsplitter was compared to a commercial beamsplitter setup (Cairn MultiCam, Cairn Research), which was used to build a second sequencing microscope. All other components of these two imaging systems are identical, allowing for a direct

comparison of beamsplitter performance. Z-stack images were taken of the same regions of tissue with both sequencing setups. The XY locations of spots (rolonies) were detected by finding the location of isolated signal peaks from the images. Cubes of pixels from the image at these locations were extracted, and the spot signal was averaged across all spots to create intensity profiles of the spot signal. The commercial beamsplitter and custom beamsplitter both performed comparably in the centre of the field of view, with a mean XY full width at half maximum (FWHM) of 0.70 µm for the commercial beamsplitter and a mean XY FWHM of 0.72 µm for the custom beamsplitter (Figure 3.4 A, B, D, G). However, at the edges of the field aberrations were more apparent in both systems. The commercial system introduced cross-shaped astigmatic aberrations (Figure 3.4 A, C), whilst the custom system introduced coma aberrations elongating point sources away from the centre of the field of view through changes in z (Figure 3.4 B, F). In both cases, these sources of aberration were minimal and only increased the XY FWHM footprint of the rolonies by a mean of 0.06 µm for the commercial beamsplitter and 0.16 µm for the custom beamsplitter (Figure 3.4 E, H).

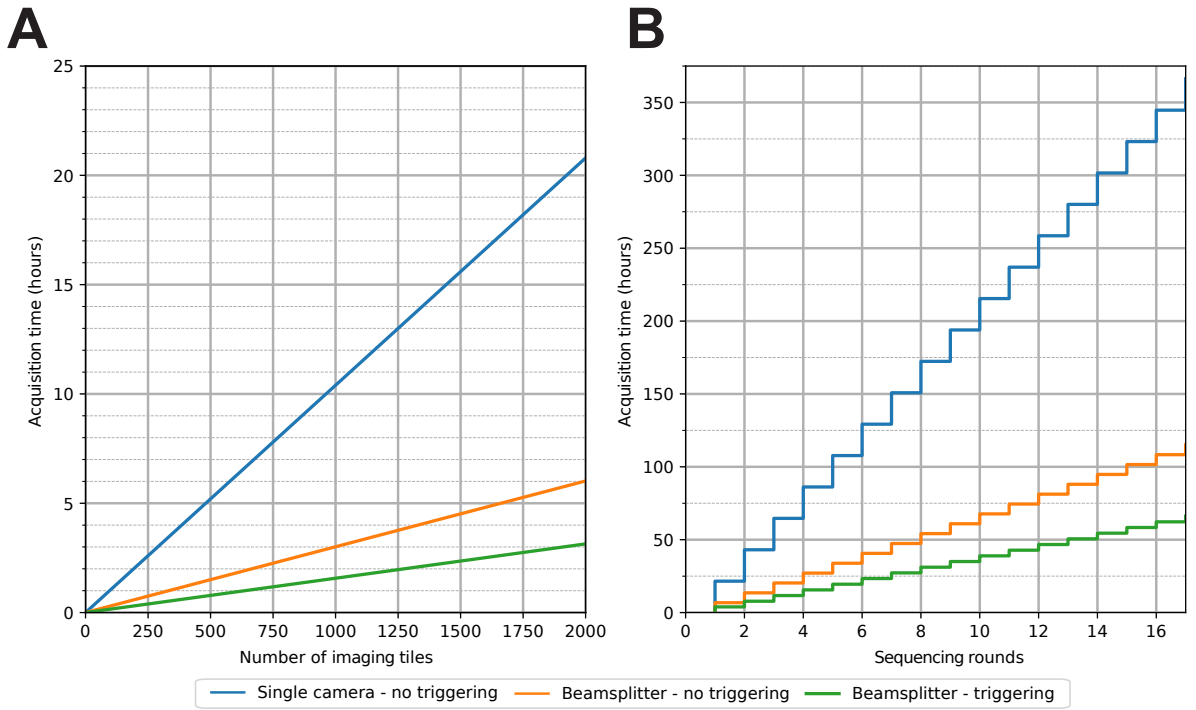


Figure 3.3: Image acquisition time is decreased by using a beamsplitter and hardware triggering of z-movements. (A) Scaling of image acquisition time for one round of sequencing with the number of image tiles acquired. **(B)** Scaling of total sequencing time with number of sequencing rounds, including image acquisition time per round and a fixed 45 minutes per round of sequencing chemistry. Blue lines represent a single camera imaging setup, yellow lines the 4-way beamsplitter without hardware triggering, and green lines the 4-way beamsplitter with hardware triggering.

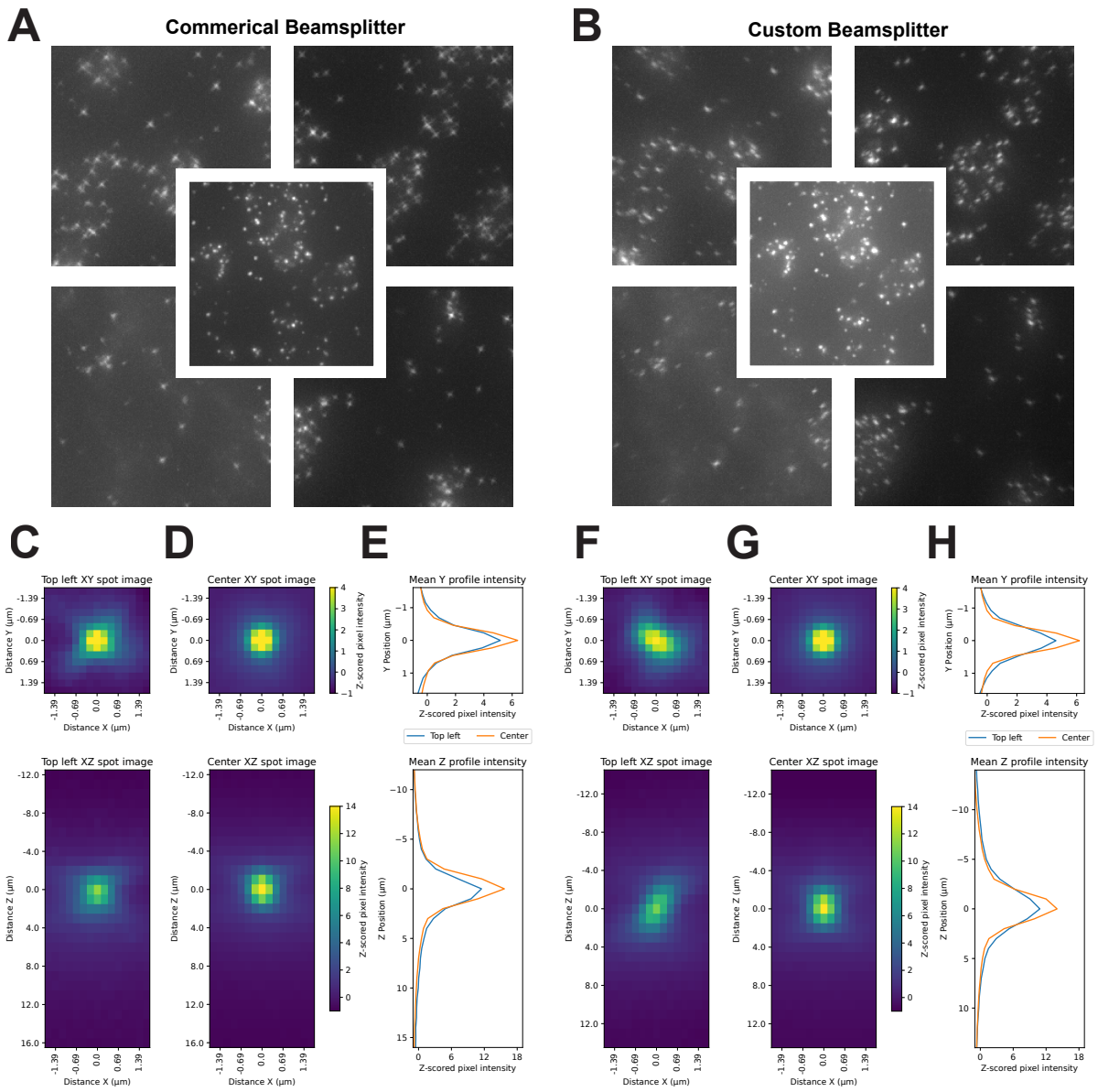


Figure 3.4: Custom beamsplitter aberrations are comparable to a commercial system. **(A)** Images from the four corners and the centre of a field of view of rolonies imaged with the Cairn commerical beamsplitter. Image sections are 200 x 200 from a total field of view of 3290 x 2460. **(B)** Images from the four corners and the centre of a field of view of rolonies imaged with the custom Thorlabs beamsplitter. Image sections are 200 x 200 from a total field of view of 3290 x 2460. **(C)** XY (top) and XZ (bottom) spot intensity profile of rolonies detected in the top left corner of the image plane **(A)** from the Cairn beamsplitter. Pixel values are the maximum z-projection (top) and y-projection (bottom) of the z-scored, mean values at each voxel of a 15x15x29 cuboid (N=113 rolonies). **(D)** As in **(C)**, but with rolonies from the centre of the image plane from the Cairn beamsplitter (N=110 rolonies). **(E)** The mean z-scored pixel intensity y-values (top) and z-values (bottom) of the spot intensity images of **(C, D)** **(F, G)** As in **(C)**, but with rolonies from the top left **(F)** and centre **(G)** of the image plane from the custom Thorlabs beamsplitter (N=141 top-left rolonies, N=144 centre rolonies) **(H)** The mean z-scored pixel intensity y-values (top) and z-values (bottom) of the spot intensity images of **(F, G)**

Comparison of vignetting

Likewise, the extent of vignetting was measured in both beamsplitters. The custom beamsplitter and the commercial beamsplitter utilised 25 mm emission filters and standard-size 25.5 x 36 mm dichroic mirrors. The custom beamsplitter was designed to mount dichroic cube holders with SM1 25.4 mm threaded ports, utilising wider SM2 50.8 mm ports and tube lenses throughout the rest of the design to reduce vignetting. When comparing the two systems, the custom beamsplitter exhibited less vignetting than the commercial design, with half the intensity drop-off at the edges of the field of view (Figure 3.5).

3.3.3 Development of a sequencing flowcell for increased tissue processing throughput and improved optical performance

Another bottleneck of sequencing throughput was the total number of tissue sections that could be processed in a single sequencing run. Originally BARseq was designed to use a 21 x 21 mm HybriWell chamber (Grace Biolabs, HBW2222FL). This is a chamber to confine reagents that can be sealed onto the surface of a microscope slide. It is composed of a

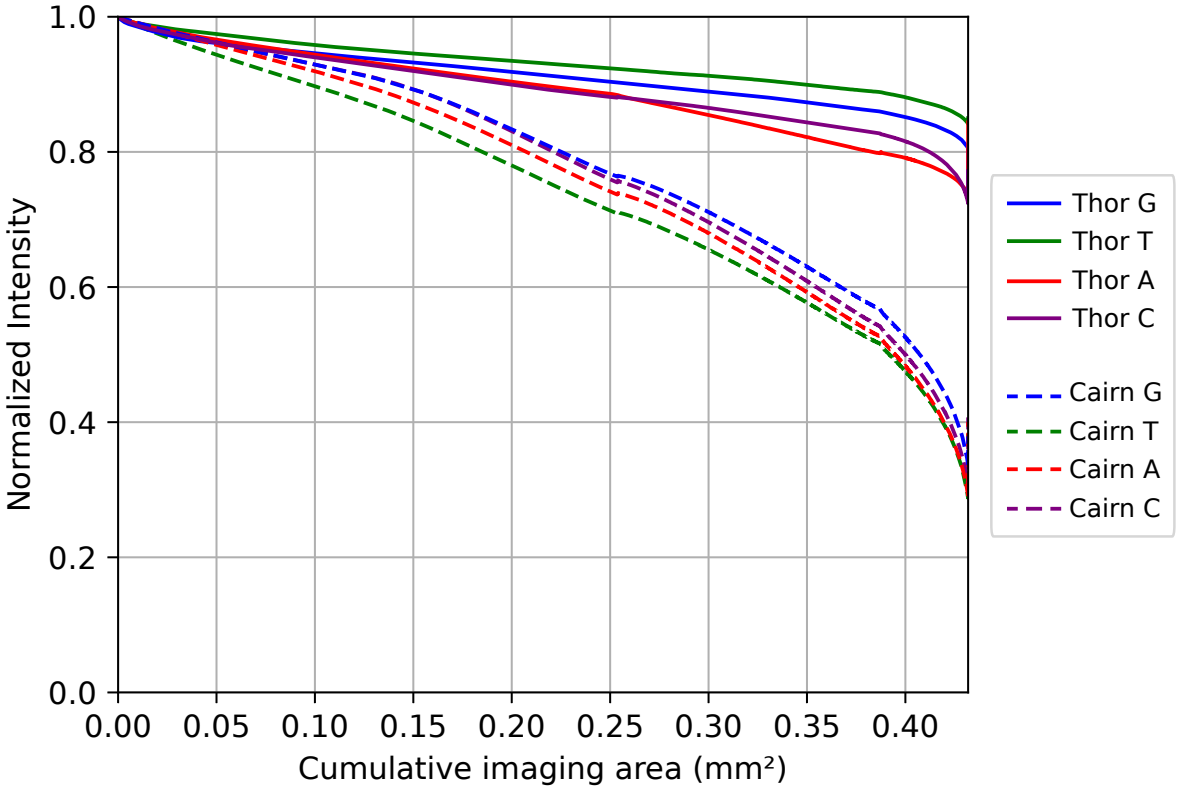


Figure 3.5: Custom beamsplitter produces less vignetting than a commercial system. Radial pixel intensity profiles of each camera FOV for the commercial Cairn beamsplitter (dashed lines) and the custom Thorlabs beamsplitter (solid lines) imaging a fluorescent acrylic slide. Intensity data is the mean of all pixels found at each distance from the centre of the 3290 x 2460 field of view, normalised to the centre intensity. G=Guanine, T=Thymine, A=Adenine, C=Cytosine beamsplitter channels

0.25 mm thick polycarbonate window with a 250 µm thick fluorescent dye-safe adhesive seal (Figure 3.6 A). This allowed for the imaging of one or two whole mouse coronal brain sections per slide with a reagent volume of 116 µl required to fill the chamber. I developed my own flowcell design to match the dimensions of a 25 x 75 mm microscope slide (Figure 3.6 B), in order to fit more brain sections onto a microscope stage than could be achieved using HybriWell chambers. Grace Biolabs adhesive material was used to form an adhesive gasket around the flow cell, laser cut to fit the dimensions of a slide whilst maximising the usable mounting area in the centre. A reduced gasket depth was used, 140 µm thick compared to the 250 µm of the HybriWell. This kept the flowcell internal volume at only 180 µl, meaning that using only 1.55X the amount of sequencing reagents, 5X the amount of tissue could be processed, decreasing cost over three-fold. Rather than mounting samples to a charged microscope slide which would have a HybriWell placed over the top of it, the samples were instead mounted directly onto 1.5H thickness coverslips, coated with poly-l-lysine to produce a charged surface for increased tissue adherence. Glass microscope slides were used to form a rigid top to the flowcell and two 1 mm diameter holes were drilled in the slides at either end to form flow ports for both manual and automated fluid transfer. These holes were covered with PDMS blocks that had 1.5 mm port holes punched into them before being plasma bonded to the glass surface to form a strong covalent bond. This allowed pipette tips or microfluidic tubing to be press-fit into the ports to form a water-tight seal for fluid flow. Four of these microscope slide-sized flowcells could be mounted onto a microscope stage for consecutive imaging (Figure 3.6 C).

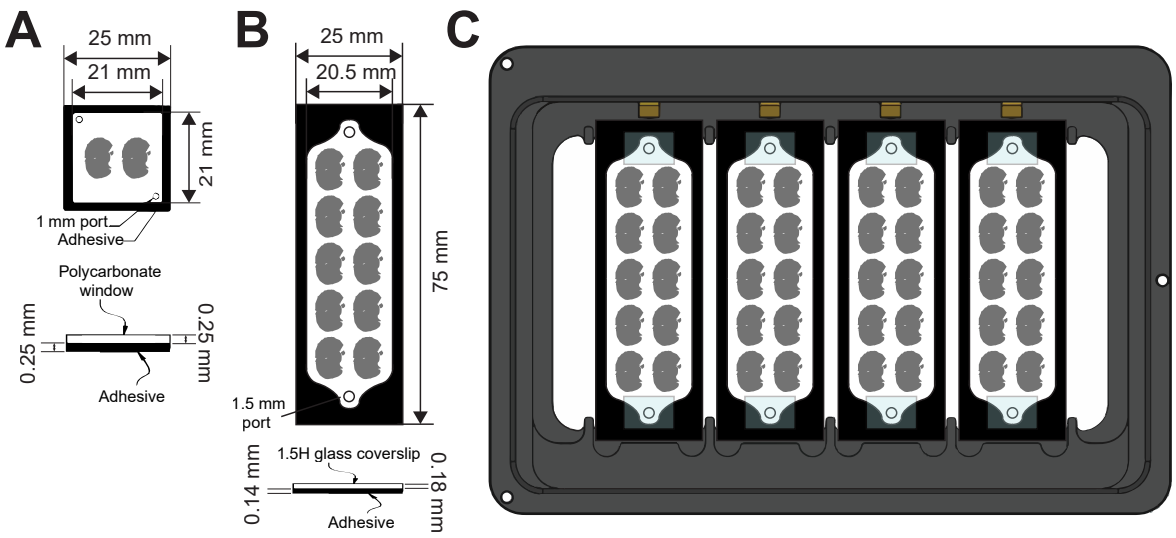


Figure 3.6: Improved flowcell design fits more tissue for increased throughput. (A) A diagram of a HybriWell chamber and its dimensions with a top-down view above and a side-on view below showing mouse coronal brain sections to scale. **(B)** A diagram of a microscope slide sized flowcell to the same scale as the HybriWell in (A) with a top-down view above and a side-on view below. **(C)** A diagram of the flowcells in the 4-slide microscope stage attachment used for the sequencing microscopes

Another advantage of the new flowcell design was the use of standardised 1.5H coverglass which microscope objectives are optimised to image through. Rather than having to pass through 250 µm of polycarbonate at the top of the HybriWell chamber and a further 250 µm of imaging medium before reaching the sample, light travels through 170 µm of glass coverslip directly to the sample, reducing spherical aberration. The extent to which this aberration was reduced was measured by comparing images of rolonies from tissue mounted in HybriWell chambers and in glass flowcells (Figure 3.7). Rolony images from HybriWells were sourced from Chen et al., 2021 BARseq imaging data (tile 20200111M1genes/M1-L_000_003/). This data was acquired with an Olympus IX81 microscope with a Crest X-light V2 spinning disk confocal, Olympus UPLFLN ×40 0.75-NA air objective with a 1 m z step size and 0.16 m xy pixel size. Images of rolonies from the glass flowcell were acquired with a Nikon Ti2-E epifluorescent microscope setup using a 20x 0.8 NA Lambda-D air objective, with a 1 µm z step size and 0.231 µm pixel size. Rolony spot signals were isolated as described for Figure 3.4 and the signal was averaged across all spots to create intensity profiles. Figure 3.7 A and D show the XY and XZ z-scored spot intensity profile of rolonies imaged through a HybriWell chamber. When compared to rolonies imaged through the custom glass flowcell (Figure 3.7 B, E), the two shared a similar lateral resolution, with a mean FWHM of 0.74 µm for the HybriWell and a mean FWHM of 0.65 µm for the flowcell (Figure 3.7 C), whilst the flowcell demonstrated an improved axial resolution compared to the HybriWell with a mean FWHM of 2.5 µm for the flow cell and 10.5 µm for the HybriWell (Figure 3.7 F). The differences in the imaging setups between the HybriWell and flowcell image acquisitions mean that this cannot treated as a direct comparison, however the confocal system used for imaging the HybriWell chambers has a theoretical lateral and axial resolution which is very similar to the microscope setup used for imaging the flowcells (HybriWell confocal microscope simulated lateral resolution: 396nm, simulated axial resolution: 1603nm, flowcell epifluorescent microscope simulated lateral resolution: 388nm, simulated axial resolution: 1590nm).

3.3.4 Development of a microfluidic system for automated in situ sequencing

The beamsplitter imaging setup enables high throughput image acquisition. However, between each imaging round, there is a necessary set of chemistry steps that have to be performed in order to cleave away the dyes and terminators of the current nucleotides before incorporating the next dye-labelled bases for imaging. This manual process involves removing the flowcells from the microscope and performing the chemistry steps by introducing a series of reagents to the flowcells via manual pipetting, using a hybridisation oven to incubate the flowcells at 60 °C between pipetting steps. Performing this manual pipetting takes about 45 minutes per round. The samples then have to be re-loaded

onto the microscope stage and offsets in X/Y/Z need to be calculated per flowcell due to inaccuracies in manual replacement of the slides round by round. This hands-on time means that consecutive rounds of sequencing cannot continue overnight without human intervention. I designed a microfluidic system that applies the reagents to the chambers automatically to try and remove this manual interaction that occurs every round (Figure 3.8).

The fluidic system is comprised of an Arduino Nano (Arduino, A000005), used to send TTL signals to a Fluigent fluidics control unit upon completion of each imaging round (Fluigent, LU-LNK-0002). Upon receiving the TTL signal, the Fluigent system engages a fluidic routine, using a pressure controller (Fluigent, ELUPPU1000) to supply a controlled pressure of compressed air to reagent vessels containing each of the necessary sequencing reagents (Fluigent, P-CAP). A 10-way rotary valve (Fluigent, ESSMSW003) selects

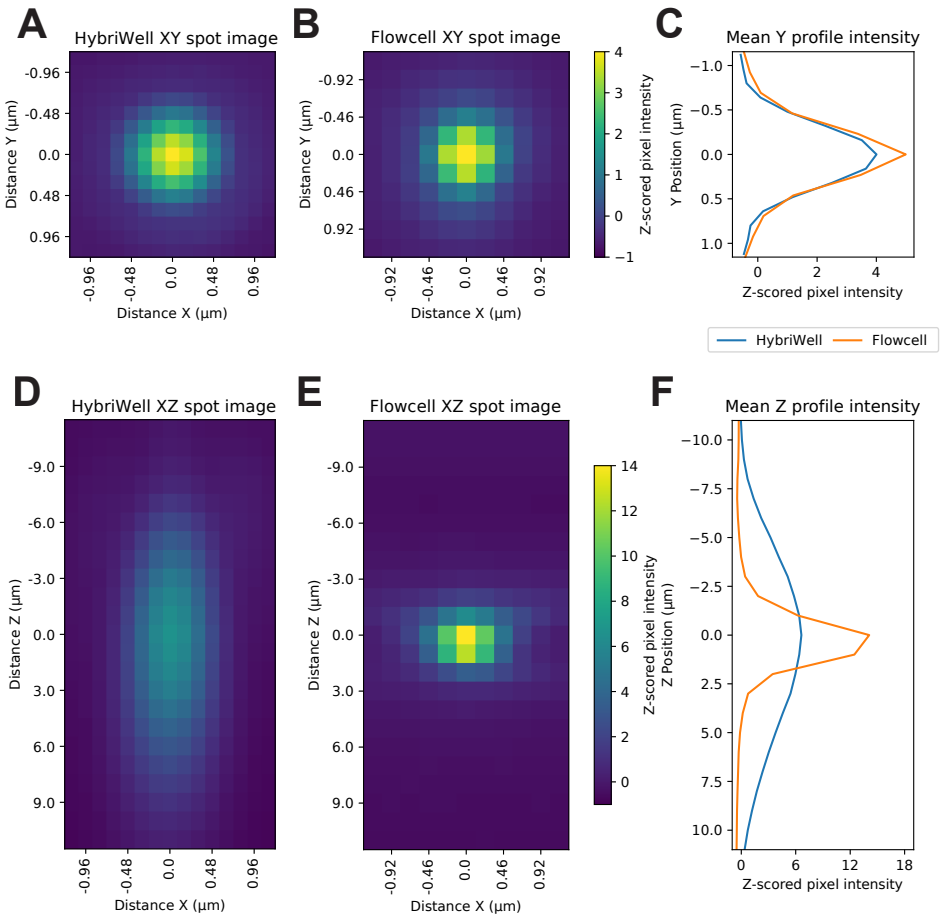


Figure 3.7: Custom flowcells reduce aberration compared to HybriWell chambers. **(A, B)** XY spot intensity profile of rolonies imaged through a HybriWell chamber **(A)**, and through a custom flowcell **(B)**. Pixel values are the maximum z-projection of the z-scored, mean values at each voxel of an 11x11x23 cuboid (N=422 rolonies) **(A)**, and a 15x15x23 cuboid (N=428 rolonies) **(B)**. **(C)** The mean z-scored pixel intensity y-values of the spot intensity images of **(A, B)**. **(D, E)** XZ spot intensity profile of rolonies imaged through a HybriWell chamber **(D)** and through a custom flowcell **(E)**. Pixel values are the maximum y-projection of the z-scored, mean values at each voxel of an 11x11x23 cuboid **(D)** and a 15x15x23 cuboid **(E)**. **(F)** The mean z-scored pixel intensity y-values of the spot intensity images of **(D, E)**.

which reagent can flow into the flowcell at each step of the program, using in-line flow meters (Fluigent, Flow M+) to measure the amount of each reagent that passes through the flowcell, providing closed-loop feedback for the pressure controller. As some of the sequencing reagents need to be kept at 4 °C to remain stable, they are stored in peltier cooling blocks (Pro-lab, PL.BSH200-E). Upon heating the cold reagents to 60 °C, microbubbles form in the flowcell, impeding fluid flow and imaging. To prevent this, an adjustable in-line heater (Francis Crick Institute Making Lab) was placed directly downstream of the rotary valve to preheat the solutions to 50 °C. A passive PTFE membrane bubble trap (Fluigent, CTQ-006BT) was added after the heater and before the flowcell to remove any bubbles from heat-induced degassing. The in-line heater and a second heater that attaches directly to the top of the flowcell are controlled by a microcontroller board using thermocouple temperature readout for closed-loop feedback, which allows for temperature set point adjustment (Francis Crick Making Lab). The flowcell is continuously maintained at 60 °C during fluidic steps with the flowcell heater. To prevent thermal expansion of the microscope optics, the sample is moved away from the objective by the microscope stage during each fluidic round. After completing the fluidic chemistry steps, the heater shuts off and the flowcell is allowed to passively cool to 25 °C before resuming imaging.

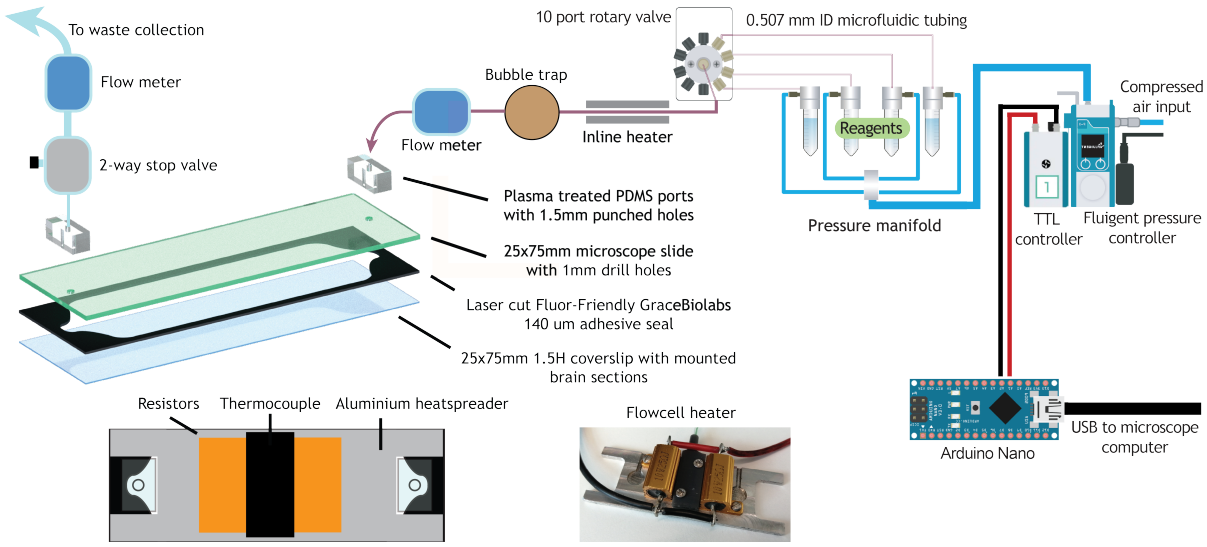


Figure 3.8: A pressure-based microfluidic system for automated sequencing reagent delivery. A diagram of the microfluidic system used to perform chemistry steps on flowcells. USB communication from the microscope PC is translated to TTL pulses that begin a fluidic routine controlled by the Fluigent pressure controller. This regulates the air pressure reaching a manifold which feeds into each reagent container. A 10-way rotary valve controls which reagent can flow into the flowcell. An inline heater preheats reagents, enabling degassing by a downstream bubble trap. Flow meters up and downstream of the flowcell monitor the flow rate to dispense the correct amount of reagent. The flowcell itself is heated by a resistive heating unit placed over the top of the glass slide of the flowcell. A 2-way stop valve prevents flow during incubation steps and leads to a waste collection receptacle.

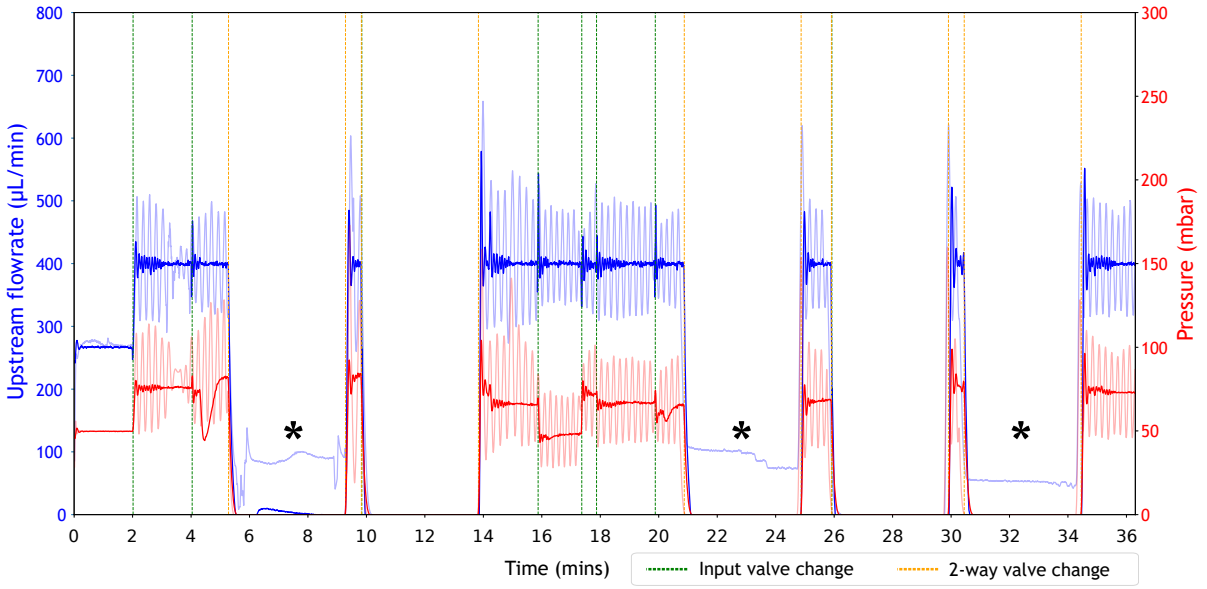


Figure 3.9: Flow rate oscillations and residual flow are solved with increased feedback response time and a stop valve. A graph showing pressure (mbar) and flow rate (µl/min) readings over the course of a sequencing fluidic program. Pale red and pale blue lines indicate initial pressure and flow rate readings, respectively. Dark red and dark blue lines indicate pressure and flow rate after increasing feedback response time from 2 to 10 seconds and adding a 2-way stop valve. Asterisks denote incubation periods with 4°C reagents entering the flowcell. Green dashed vertical lines indicate time points when the 10-way valve changed reagents, and yellow dashed lines indicate time points when the 2-way stop valve was opened or closed.

Optimising the control of microfluidic flow rate

I observed a "ringing" like oscillation in flow rate when configuring the closed-loop feedback from the flow rate sensors to the pressure controllers (Figure 3.9). As there was a delay from applying pressure to the compressible air in the manifold to observing a flow rate change downstream, this oscillation would not decrease even over time frames longer than 5 minutes. To improve the reproducibility of the fluid delivery, I increased the response time over which the feedback occurred from 2 to 10 seconds, which damped the oscillation considerably, with decay to a steady flow rate observed in around a minute after changing the flow rate set point. Before an in-line heater and 2-way stop valve were added, another inconsistency was observed in the fluid flow. When the flow was halted by decreasing pressure to 0 mbar, a residual flow rate of up to 100 µl/min was observed when a cold 4 °C reagent was pumped into the flowcell (asterisks in Figure 3.9). The flowcell heating unit was constantly set to 60°C, so thermal expansion of the fluid resulted in a continued flow during incubation. To prevent this from occurring, a 2-way stop valve was installed downstream of the flowcell, which would shut off the waste outlet during incubation steps and prevent fluid movement. The upstream in-line heater was installed to degas dissolved CO_2 from the cold reagents, but also mitigated the thermal expansion of the fluids entering the flowcell by preheating them to near the flowcell set temperature.

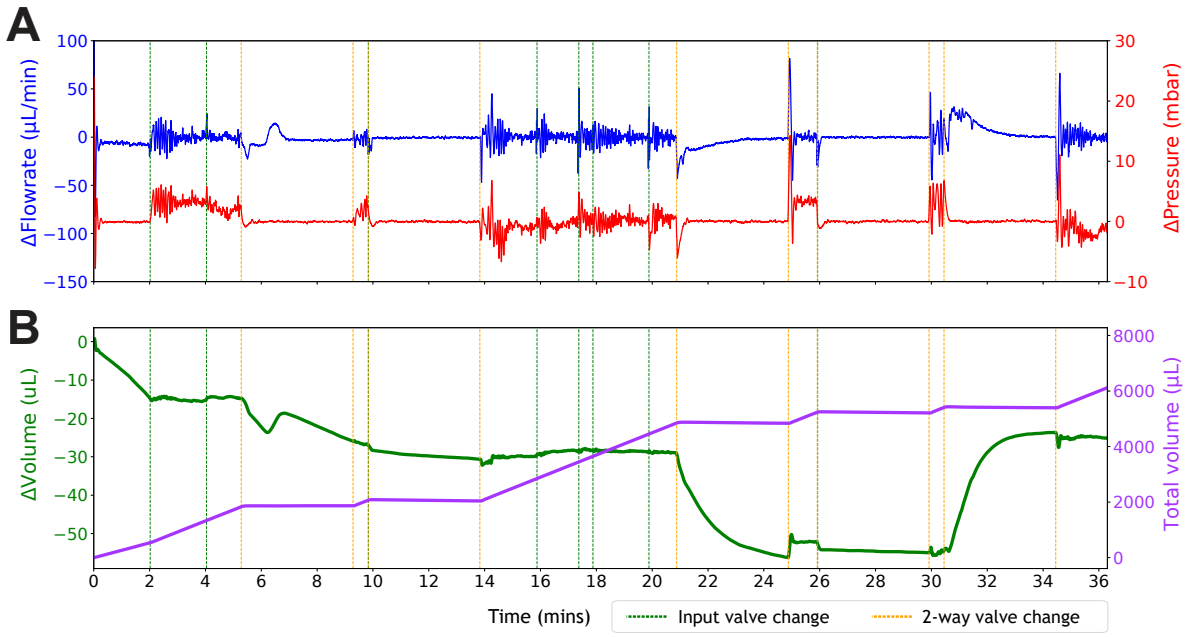


Figure 3.10: The pressure-based microfluidic system delivers reproducible volume across rounds. **(A)** A graph showing pressure (mbar, red line) and flow rate (µl/min, blue line) differences between two fluidic program runs. **(B)** A graph with the same time scale as **(A)** showing the cumulative change in volume over time (µl) between the two fluidic program runs in green and the total volume of flow (µl) from the second fluidic program run in purple. In both plots, green dashed vertical lines indicate timepoints when the 10-way valve changed reagents, and yellow dashed lines indicate timepoints when the 2-way stop valve was opened or closed.

These steps to improve the reproducibility of the fluid volume delivered to the flowcell during the sequencing program were then quantified by performing repeated runs of the same fluidic program. The differences in the pressure applied and flow rate observed in the system were determined by first smoothing the readout with a sliding window of 2 seconds (sensor readout was acquired every 100 ms) and subtracting the values of one round from the other (Figure 3.10 A). From the flow rate difference, the cumulative flow difference and total observed flow volume from the second fluidics run were calculated (Figure 3.10 B). Over the course of a 36-minute fluidic program with 10 reagent changes and a total flow volume of 6118 µl, a peak volume difference of -56.2 µl was observed (0.92 % of the total volume delivered), most of which was due to differences in residual flow during later incubation steps rather than amounts dispensed during reagent delivery.

Automated microfluidic sequencing failure modes

Despite the consistent flow rate that the microfluidic system can achieve, automated sequencing results were variable. Initial microfluidic testing was performed with the flowcell heating unit placed in direct contact with the underside of the tissue-containing coverslip to ensure the tissue reached the target temperature. The final imaging configuration requires the heater to be placed on top of the flowcell, to allow optical access to the tissue.

This requires heating the sample through the flowcell's top microscope slide glass layer and the fluid medium, raising the possibility of inconsistent sample heating. Successful rounds of gene sequencing were achieved with automated microfluidics when the flowcell heater was mounted directly to the coverslip (Figure 3.11 A). However, in sequencing attempts when the heater was placed on top of the flowcell, failure modes occurred with rolonies losing signal rapidly in the first few rounds (Figure 3.11 B), or with tissue background increasing rapidly in the channel of the cytosine conjugated dye (that of the longest emission wavelength fluorophore at 700 nm+) (Figure 3.11 C). In the case of rolony signal loss, this occurred when the set point temperature of the heater was increased to 70 °C to try and compensate for inefficient heat transfer to the sample, likely resulting in signal loss due to the melting of the sequencing primer and its extended bases from the barcodes of the rolonies (gene sequencing primer PAB0018 used in these experiments had a calculated Tm of 63.4 °C). Setting the flowcell heater temperature to a low temperature (37 °C) recapitulated the increase in background fluorescence build-up. This background primarily accumulated in the white matter tracts of the brain and the membranes of cell bodies (Figure 3.11 D). Increasing the number of PBST washes post-nucleotide incorporation decreased this tissue background, as did increasing the percentage of Tween from 0.5% to 2% in the PBST as has been used in updated BARseq protocols (Chen, Fischer, et al., 2024; Zhang et al., 2024), but this did not fully abolish the background signal. After multiple rounds of sequencing, this background exceeds the signal of the rolonies and can prevent barcode readout. The performance of the flowcell heater was exacerbated by a curvature of its aluminium heat spreader, causing an imperfect interface with the glass surface of the slide and by drifting readings from the thermocouple used to maintain the temperature set point. However, this background buildup also occurred during manually pipetted sequencing in some samples, where heating should be consistent due to the use of a hybridisation oven and metal sample blocks for maintaining the 60 °C target temperature. It is also possible that the slow flow rate of the microfluidic system (400 µl/min) did not induce sufficient reagent mixing at the surface of the tissue within the flow chamber due to frictional boundary effects.

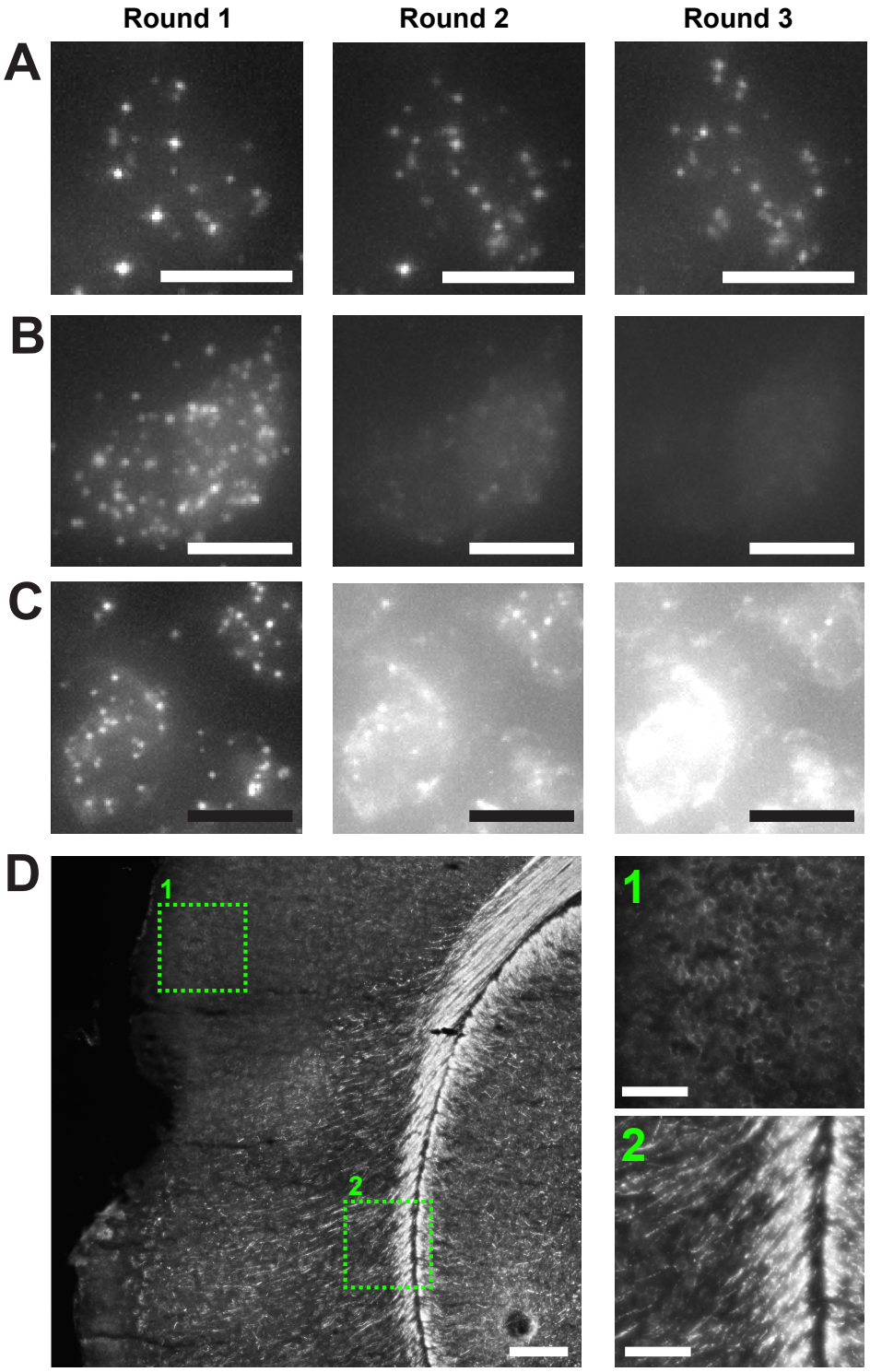


Figure 3.11: Variability in automated sequencing signal. (A, B, C) Example images of cells with rolony fluorescence signal from the cytosine dye channel. The first three imaging rounds from gene barcode sequencing are depicted. Scale bar = 10 µm **(A)** A successful automated sequencing run with flowcell heater in direct contact with the coverslip. **(B)** Automated sequencing performed with the heater on top of the flowcell, set to 70 °C. **(C)** Automated sequencing performed with the heater on top of the flowcell, set to 60 °C with background fluorescence buildup. **(D)** An example image of mouse cortex after one round of automated sequencing with the heater on top of the flowcell set to 37 °C (Scale bar = 200 µm), insets show 1) cell membrane background and 2) white matter background build-up. Scale bar = 50 µm

3.4 Discussion

The ability to read random sequences of bases from nucleic barcodes and attribute them with confidence to individual cells is crucial for barcoded connectomics but is hard to achieve at scale. Out of the in-situ sequencing techniques that have been described to date, BARseq is uniquely suited for this task as a method which can achieve subcellular spatial resolution of transcripts whilst determining arbitrary, unknown sequences. As this method requires thin tissue sections, it is necessary to sequence many of these in order to cover a brain volume of interest. Here, I made several advancements to increase the throughput at which BARseq in situ sequencing experiments can be performed. As imaging time was the longest component of these experiments, I focused on reducing this time by acquiring all four nucleotide dye channels at once with a custom beamsplitter setup. I likewise expanded the amount of tissue that could be processed at once with custom flowcells that fit more tissue whilst using less reagent per tissue section. Finally, I produced a microfluidics system for automating the sequencing process from start to end by integrating fluidic control of the reagents required for the sequencing chemistry with the software based control of the microscope. This fluidic system could deliver reagents consistently to a single flowcell, but problems with maintaining a consistent temperature across the flowcell, along with the high levels of maintenance required to purge the system of air to keep flow rate recordings from drifting, meant that the experiments carried out in later chapters did not rely on this fluidics system. Efforts to continue my full automation of BARseq are being continued in collaboration with the Francis Crick Making Lab to improve the flowcell heating units and use a simpler volume-defined, syringe-based microfluidic setup which would be able to flow reagents through all four flowcells during a sequencing experiment. The improvements to imaging that I have already outlined here allow for 800 µm of serial tissue sections to have their endogenous gene transcript 7 bp barcodes sequenced within a week, and for a combined endogenous gene transcriptomics and barcoded rabies connectomics experiment to be carried out on a brain in under a month.

Chapter 4

Identifying transcriptomic cell types with in situ sequencing

4.1 Authorship declaration

I, Alexander Becalick, performed all of the experiments and the majority of the analyses in this chapter with the exceptions indicated here. Antonin Blot performed the surgical procedures on the mice. Petr Znamenskiy, Antonin Blot and I collaboratively developed the data preprocessing pipeline for detecting genes and assigning them to cells.

4.2 Introduction

In a rabies tracing experiment, information on the connectivity of neurons is provided by the locations of both the input cells and the post-synaptic cells that they make connections with. This information can already provide an understanding of how feedforward, feedback, or recurrent loops of connectivity allow the processing of information between different regions and layers of the cortex and beyond (Graf et al., 2002; Kelly and Strick, 2003, 2004; Ohara, 2009; Ugolini, 1995). However, without additional context, connectivity information has limitations. Neurons in the brain are highly diverse, with distinct molecular, electrophysiological, and morphological properties that define their roles within circuits. Connectivity by itself does not provide insight into the roles of the neurons involved in these circuits, whether they are specialized for specific tasks or even whether they are excitatory or inhibitory. Previous studies using rabies viruses to study connectivity have used pseudotyped rabies viruses along with cell type-specific Cre/Flp/tTA driver mouse lines and recombinase/driver-dependent helper viruses to selectively infect defined starter cell types (Hafner et al., 2019; Li et al., 2013; Miyamichi et al., 2011; Wall et al., 2010; Weible et al., 2010). This strategy enables studying the input onto specific cell types and can be extended further through the use of immunohistochemistry

(Stepien et al., 2010), in-situ hybridisation (Weissbourd et al., 2014) or single-cell RNA sequencing (Patiño et al., 2024) to study presynaptic cell types. By choosing to use an in situ sequencing technique to determine the barcode sequences of rabies-infected cells, I have selected a method which simplifies the concurrent determination of endogenous transcripts. Using the same spatial transcriptomic modality, both types of information can be extracted from the tissue in the same experiment.

Over the past two decades, the ability to study the transcriptome of individual neurons in mammalian brains has rapidly expanded in scope, resulting in datasets that first contained hundreds, then thousands and recently millions of cells due to advancements in single-cell and single-nuclear RNA sequencing. These efforts began as attempts to provide deeper insight into the diversity of specific groups of neurons (Chen et al., 2017; Li et al., 2016; Poulin et al., 2014) or as attempts to broadly sample neurons of many types, albeit sparsely (Darmanis et al., 2015; Lake et al., 2016; Zeisel et al., 2015). With the establishment of the BRAIN Initiative Cell Census Network (BICCN), studies have sought to produce a comprehensive transcriptomic cell-type atlas of the entire mouse and human brain, an effort that has resulted in deep coverage of all mouse anatomical regions through scRNAseq sampling of hundreds of microdissected brain regions (Yao, van Velthoven, et al., 2023), along with companion spatial transcriptomic data matched to the same cell type clusters (Zhang et al., 2023). This comprehensive atlas, along with more specialised taxonomies that came before it to describe the visual cortex in detail (Tasic et al., 2016, 2018) can be used as ground truth transcriptomic references when planning experiments to separate observed neurons into cell types based on the expression of a limited set of marker genes.

As described in Chapter 3 of this thesis, I employed BARseq to perform in situ sequencing of both rabies barcode transcripts and endogenous gene transcripts. BARseq has previously been demonstrated to enable cell-type classification in the mouse motor cortex and auditory cortex of cells infected with barcoded Sindbis viruses (Sun et al., 2021) and more recently has been used to study brain-wide cortical modules of transcriptomic cell types in the mouse cortex (Chen, Fischer, et al., 2024) and to determine the cell types of neurons that have been barcoded by rabies viruses (Zhang et al., 2024). I aimed to design a panel of gene padlock probes which maximises the accuracy of neuronal cell type assignment in the mouse visual cortex. I then aimed to test both the fidelity of the individual gene probes and the accuracy of the clustering performed on cells using these probes, by using the spatial information provided by BARseq to validate gene expression patterns and cell type locations against known data from the mouse visual cortex.

4.3 Methods

4.3.1 Gene selection and padlock probe design

Gene panel selection for cell type classification

SMART-Seq2 RNA expression data from 13,465 cells from the mouse visual cortex (Tasic et al., 2018) was used as a reference for gene selection. Firstly, genes from the SMARTseq dataset were filtered to remove those with low mean expression which were unlikely to be detected with the limited sensitivity of BARseq. Mean gene expression was calculated per annotated cell type cluster and genes with <500 mean exon reads per cell in the highest expressing cluster were removed. Then, 140 candidate gene lists were generated from the remaining genes by applying a matrix of thresholds varying across two more thresholding parameters, median gene expression and gene expression skewness. Skewness was calculated from the mean gene expression values per cluster as the adjusted Fisher-Pearson standardized moment coefficient of skewness:

$$G_1 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Where n is the number of observations, $\bar{x}$ is the sample mean, and s is the sample standard deviation. Searching through each of these filtered gene lists, I used a modified version of a greedy combinatorial search algorithm (Qian et al., 2020) to iteratively select the top 80 genes that maximise classification accuracy across all cell type clusters. The algorithm works as follows: first, to account for the lower detection efficiency of BARseq compared to SMARTseq, the cluster-mean values and individual cell count values are resampled from a binomial distribution with an estimated efficiency of 0.001 (estimated by dividing the observed mean in situ sequencing read counts per cell from Bugeon et al., 2022 with the mean gene count (exon reads) per cell of the same subset of marker genes in the reference SMARTseq dataset (Tasic et al., 2018)). Using parameters of the negative binomial distribution, the log-likelihood of the observed expression counts for every cell/gene combination is precomputed for each cluster, given that cluster's mean expression value per gene. The algorithm runs for a specified number of iterations (100), starting with an empty gene set. In each iteration, the greedy algorithm searches for the gene that, if added to the current gene set, maximizes classification accuracy. Accuracy is defined as the percentage of cells assigned to their true clusters based on the currently selected gene(s), using the sum of the precomputed log negative binomial probabilities to assign each cell to its most likely cluster. The algorithm adds the gene that results in the greatest accuracy improvement to the gene set. As the gene set increases in size, genes will individually confer less information as they begin to contain redundant cluster information. At each iteration, the algorithm checks if removing any of the currently

selected genes would further improve classification accuracy in the context of the new genes. If so, it removes that gene from the set to refine the selection further. The top 80 genes were selected from the candidate panel which maximised classification accuracy whilst minimising the total median expression of the whole gene panel to avoid optical crowding from selecting too many highly expressed transcripts.

Additional marker genes

Sixteen genes were then added manually to the 80 genes that were selected with the greedy algorithm. These were selected for one of three reasons, firstly to add in canonical marker genes from scRNAseq literature in the visual cortex (*Chodl*, *Cux2*, *Fezf2*, *Foxp2*, *Rorb*, *Vip*), secondly to add marker genes that were found to be differentially expressed in rabies-infected neurons from cell types where the canonical marker is downregulated by rabies infection (Patiño et al., 2022):

Cell subclass	**Traditional marker**	**Rabies infected marker(s)**
L5 NP	*Lypd1*	*Tshz2*
Sst	*Sst*	*Grin3a / Reln*

and thirdly to specifically improve the classification of neurons within clusters (L2/3 IT neurons - *Rrad*), or between subclasses:

Cell subclass	**Extra marker gene(s)**
L2/3 IT	*Rrad*
L5 IT/PT/NP	*Etv1*, *Tle4*
L6 IT/CT	*Rasl10a*, *Hs3st4*
Vip/Sst	*Grik1*, *Sorcs3*

Gene panel classification accuracy assessment

To assess the robustness and generalisability of cluster assignment accuracy for my final gene panel, I employed a repeated random subsampling cross-validation approach. Specifically, I randomly partitioned the data such that 66% of the cells from each cluster were used as a training set from which I computed the cluster-specific mean gene expression values for the selected genes. The remaining 33% of the cells served as an independent test set. These held-out cells were classified using the gene panel's cluster means derived from the training subset, and classification accuracy was recorded. I repeated this partitioning and evaluation process 100 times, each time producing a new random split and an independent accuracy estimate. The final reported accuracy and its variability were calculated by averaging across all iterations, providing a more robust estimate of the panel's classification performance compared to relying on a single train-test division.

Padlock probe design

Multiple padlock probes were designed per gene to increase detection efficiency, and low-expressing genes had more probes designed against them. The number of probes to select for a given gene was calculated using the highly expressed *Sst* gene as a reference, for which 4 padlock probes were designed:

$$N = \left(\frac{X_{Sst}}{X_{target}} \right) \times 4$$

where N is the number of padlock probes, X_{Sst} is the mean expression of the *Sst* gene in the *Sst* cell type subclass, and X_{target} is the mean expression of the target gene in its highest expressing cluster. This was limited to an upper bound of 12 padlocks per gene and a lower bound of 4 padlocks per gene.

Padlocks were generated using a modified version of the 'Multi padlock design' tool (https://github.com/Moldia/multi_padlock_design). Briefly, for each target gene, the tool finds the consensus sequence from all transcript variants in the mouse RefSeq transcriptome (https://ftp.ncbi.nlm.nih.gov/refseq/M_musculus/mRNA_Prot/) and tests every possible candidate sequence in a sliding window along the transcript. Candidate sequences are the length of the two adjacent padlock arm regions (20 bp per arm). The arms of each candidate sequence are filtered to be within a defined melting temperature range (60 - 75°C) after formamide-containing hybridisation buffer correction (melting temperature calculated assuming 0.1 µM probe, 0.075 M Na^+, 0.01 M Mg^{2+} and 20% formamide). The specificity of each candidate sequence is then tested with a BLAST search against the whole mouse transcriptome to ensure padlocks do not exhibit non-specific binding. A minimum base pair distance between padlock candidates is also specified (4 bp) to avoid steric hindrance of bound padlocks.

All padlocks have a conserved backbone sequence which contains a shared sequencing primer binding site directly upstream of a 7bp gene barcode, unique to each gene. Gene barcodes were selected from a pool of previously designed "GII codes" (Sun et al., 2021) with a minimum hamming distance of 2 between each barcode. I designed an extended set of 10 bp barcodes for future gene panels with 1432 barcodes. In addition, these have a minimum hamming distance of 4. Padlock probes were designed for the highly expressed genes Slc17a7, Gad1, Sst and Vip with unique binding sites in place of the sequencing primer site to allow for direct readout of these genes with fluorophore-conjugated hybridisation probes separately from the sequencing rounds of the other genes.

4.3.2 Animals and surgical procedures:

All animal procedures were conducted in compliance with the guidance and regulations set forth by the UK Home Office according to the Animals Scientific Procedures Act 1986.

The surgeries performed were approved by the Named Veterinary Surgeon at the Francis Crick Institute under project licence PP4882546. Wild-type male and female C57BL/6 mice aged between 12 and 15 weeks were used for gene panel transcriptomic experiments. To extract the brains for in situ sequencing, mice were euthanised with an IP injection of sodium pentobarbital. The brains were extracted without prior perfusion and flash-frozen in OCT-filled cryomoulds using a bath of absolute ethanol and dry ice, and subsequently stored at -80°C.

4.3.3 In situ sequencing methods

Flowcell fabrication:

25 x 75 mm uncoated coverslips (ibidi GmbH, 10813) were rinsed in water and then coated for 5 minutes with poly-l-lysine solution (Sigma Aldrich, P8920) diluted to a concentration of 0.01% in PBS. Slides were rinsed again in water to remove residual PBS and dried overnight before use in cryosectioning. PDMS blocks were cast by mixing Sylgard 184 silicone elastomer (Dow, 1673921) curing agent and base in a 1:10 ratio by weight in a Thinky Mixer (Intertronics, THI ARE-250) with the following protocol: 5 minutes mixing at 2000 RPM, 1 minute defoaming, 7 minutes mixing at 2200 RPM, 2 minutes of defoaming. PDMS was then further degassed in a vacuum chamber for 30 minutes before use. PDMS was poured into an aluminium pie dish with a silicon wafer placed in the bottom to a depth of 0.5-1cm. The PDMS was baked at 110°C for 15 minutes and then removed from the wafer. Scalpels were used to cut the block into 1 cm2 pieces which were then hole-punched with a 1.5 mm biopsy punch to create flow ports for inserting pipette tips/microfluidic tubing into. 1 mm diameter holes were drilled at either end of unfrosted 25 x 75 mm glass microscope slides (VWR International, 631-1550) using a 1 mm diamond-tipped rotary hand tool bit (Drilax, 600095802246) submerged in water. Both the glass slides and PDMS ports were cleaned with ethanol then rinsed in ddH2O and dried with filtered compressed air. The surfaces of the glass slides and PDMS ports were then functionalised by exposing them to an oxygen plasma using a plasma oven (Henniker Plasma, HPT-100). Slides and ports were exposed to plasma for 30 seconds at 50 % power in 7 sccm of oxygen. The surfaces were then immediately contacted, aligning the drilled slide holes with the punched PDMS ports and baked at 80C for at least 1 hour.

Tissue treatment:

Brain cryosectioning The following protocol for amplifying transcripts from tissue was modified from BARseq2 and coppaFISH protocols (Sun et al., 2021; Bugeon et al., 2022). Brain tissue was first flash-frozen by embedding in optimal cutting temperature compound (OCT) filled cryomolds. Serial sections were cut at 20 µm thickness with MX35

Ultra disposable microtome blades (Epredia, 3053835) at -13°C specimen temperature / -15°C chamber temperature using a Bright OTF5000 cryostat (Bright Instruments, R080) and mounted onto poly-l-lysine coated 25 x 75 mm coverslips. Ten sections were mounted per coverslip and then stored at -80°C.

Buffer preparation For the following steps, all solutions were prepared from RNAse-free water. Solutions:

- PBS - 10X PBS solution pH 7.4 (Scientific Laboratory Supplies, NAT1006), RNase-free water (VWR International, VQ65052.1000)
- 4% PFA in PBS - 10 ml 16% PFA, methanol-free (Thermo Fisher Scientific, 28906), 30 ml 1X PBS
- PBST 0.5% Tween - 50 ml 1X PBS, 0.25 ml Tween20 (Sigma-Aldrich, P9416)
- Hybridisation mix - 10% Formamide (Merck Millipore, 344206), in 2X SSC buffer (Thermo Fisher Scientific, AM9763)

Flowcell assembly Coverslips were immersed in a freshly made fixative solution of 4% PFA in PBS for 30 minutes. The coverslips were then immersed in PBS to wash and then immersed twice in PBST. Next, they were immersed in a series of ethanol solutions, incubating for 5 minutes each in 70%, 85% and 100% ethanol. The coverslips were transferred to fresh falcon tubes containing 100% ethanol and stored overnight at 4°C. The following day, coverslips were removed from the ethanol, and once dried the flowcell was assembled by pressing the coverslip onto the exposed adhesive gasket of the slide. Assembled flowcells have a volume of 180 µl and were washed with PBST three times to rehydrate the tissue and remove any bubbles from within the flowcell.

Reverse transcription A reverse transcription solution was then added to each flow cell. The mix contained: pooled gene-specific reverse transcription primers (60nM per primer, 1 U/µl RiboLock RNase Inhibitor (Thermo Fisher Scientific, EO0381), 0.2 µg/µl non-acetylated BSA (Thermo Fisher Scientific, AM2616), 0.5 mM dNTP (Thermo Fisher Scientific, R0192), 20 U/µl SuperScript IV reverse transcriptase (Thermo Fisher Scientific, 18090200), 5 mM DTT (Thermo Scientific, 707265) in 1X SuperScript IV buffer. Samples were incubated at 45°C overnight in a humidified chamber to reduce evaporation.

Padlock ligation The following day, samples were washed in PBST once and the cDNA from reverse transcription was crosslinked by incubating the samples in 4% PFA in PBS for 30 minutes at room temperature. Samples were then washed in PBST twice to remove excess PFA. RNA template digestion, padlock hybridisation to cDNA and padlock ligation

were all carried out in one reaction. This gene padlock ligation mix contained: pooled gene padlocks (60 nM of each padlock), 0.4U/µl RNaseH (New England Biolabs, M0297L), 20% Formamide, 50 mM KCl, 0.5U/µl Ampligase (Lucigen, A0102K) in 1X Ampligase buffer. Samples were incubated for 40 minutes at 37°C, followed by 1 hour at 45°C.

Rolling circle amplification An RCA primer mix (PAB0010, 12, 13, 22, 33, 35, Appendix Table 7.4) was then hybridised at room temperature for 15 minutes. The samples were washed two times in hybridisation mix for 2 minutes each and once in PBST. A rolling circle amplification mix was then made, containing: 0.25 µg/µl non-acetylated BSA (Thermo Fisher Scientific, AM2616), 250 µM dNTP (Thermo Fisher Scientific, R0192), 125 µM 5-(3-Aminoallyl)-dUTP (Thermo Fisher Scientific, AM8439), 5% glycerol, 0.2U/µl EquiPhi29 DNA polymerase (Thermo Fisher Scientific, A39391) in 1X EquiPhi29 buffer. The RCA mix was added to the samples and incubated overnight at 30°C in humidified chambers. The following day, samples were washed in PBST once, then crosslinked in 50 mM BS(PEG)9 (Thermo Fisher Scientific, 21582) in PBS for 1 hour. Excess BS(PEG)9 was quenched by washing the samples twice with 1 M Tris-HCl pH 8.0 buffer and then incubating at room temperature for 30 minutes. The samples were then washed in PBST.

Microscopy

Samples were imaged using the custom-built sequencing microscope described in Chapter 3.3.1. All tissue imaging was performed using a 20x 0.8 NA Lambda-D air objective, with a CoolLED pe-4000 light source. Images were captured with 0.231 µm pixel size in 23 plane Z-stacks with a Z spacing of 1 µm and tile overlap of 10%.

Sequencing and imaging workflow

After rolony production was complete, a gene sequencing primer was hybridised to the rolonies of the samples. The samples were first equilibrated in hybridisation mix, then 1 µM LNA gene sequencing primer (PAB0032, Appendix Table 7.4) in hybridisation mix was added to the sample and incubated at room temperature for 1 hour. Excess sequencing primer was removed with three washes of hybridisation mix at room temperature for two minutes each. Then samples were washed twice in PBST before proceeding to sequencing. Sequencing was performed manually by pipetting Illumina reagents into the flowcells, placing them in a hybridisation oven for incubation at 60°C and then using a vacuum pump to remove fluids before the next step of pipetting. The samples were cooled to room temperature each round before being imaged to prevent fluctuations in focus from the infrared-based "Perfect Focus System" of the microscope. Sequencing reagents were sourced from HiSeq SBS v4 kits (Illumina, FC-410-1003) or MiSeq Reagent

v3 kits (Illumina, MS-102-3003). The following protocol was used for the first sequencing round:

1. Add 200 µl of HB2 and incubate at 60°C for 3 minutes, then repeat.
2. Wash with 200 µl Iodacetamide.
3. Add 200 µl Iodoacetamide (9.3mg in 2 ml PBST) and incubate at 60°C for 3 minutes, then repeat.
4. Wash the sample with 200 µl PBST
5. Wash the sample with 200 µl of HB2, then repeat.
6. Add 200 µl of IRM and incubate at 60°C for 3 minutes, then repeat.
7. Wash the sample with 200 µl PBST
8. Wash the sample with 200 µl PBST and incubate at 60°C for 3 minutes, then repeat three more times.
9. Add 200 µl USM and image.

After sequencing round one, the following steps were performed followed by the above procedure:

1. Wash the sample with 200 µl HB2, then repeat two times.
2. Add 200 µl CRM and incubate at 60°C for 3 minutes, then repeat.

Seven rounds of sequencing were performed to read out the full gene padlock barcode.

Hybridisation rounds After these rounds of sequencing, the gene sequencing primer and its extended barcode sequence were stripped from the sample to prepare for hybridisation probe imaging. The samples were again equilibrated in the hybridisation mix, and then 1 µM each of a set of hybridisation probes with 5' conjugated fluorophores were hybridised to the samples in two extra rounds. Each round followed the same hybridisation and stripping procedure as for the sequencing primers. Firstly, probes designed against the backbone of highly-expressed gene padlocks: *Slc17a7*, *Sst*, *Gad1* and *Vip*, were bound to the samples. These were imaged and then stripped using a solution of 60% formamide in 2X SSC buffer. Then hybridisation probes that targeted a shared sequence on the backbone of all gene padlocks and all rabies barcode padlocks were bound. After adding these 'anchor' probes, the tissue was stained with 5 µg/µl DAPI diluted in PBST for 10 minutes, followed by three washes in PBST before proceeding to imaging.

4.3.4 Image processing and gene barcode detection:

Image projection All tiles from each round of sequencing are first z-projected, taking the maximum and median intensity values from the raw image z-stack and these projected images are saved and used for all subsequent analyses. Gene barcodes are detected from maximum intensity z-projections of the raw imaging data with median intensity z-projections subtracted. A normalised average tile image is created for each round of imaging by computing the mean intensity across all tiles, which is used for illumination correction. The z-projected 2D images are then registered together across rounds and between channels.

Image registration between channels and rounds A reference tile is selected for initial registration. First, the images across sequencing rounds for each channel are all registered to the image from the first round. Images are median-filtered to reduce noise and then the rotation and x/y translation values are estimated for each image by performing phase correlation between the target and reference. The optimal rotation angle is searched for over multiple iterations, starting with a wide search range and refining the angle range by a factor of 5 per iteration. Once all angle and shift values across rounds have been calculated, the standard deviation projection of the round images is calculated, generating a reference image for each channel. These channel images are then registered together. Due to aberrations within the beamsplitter design (see Chapter 3), it is necessary to calculate the affine transformation between these images to ensure rolonies in the corners of the images remain aligned between channels. This is done by splitting each channel image into small blocks, from which the x/y shifts are calculated between each block in the target and reference image. Then the resultant shifts are fit across x and y separately with a Huber regressor, which is robust to outlier values. The regression coefficients and intercept from the x and y regression are combined to form the 6-parameter affine transform which can be applied to the images to register across channels. This set of pre-computed transformations on the reference tile is then used as the initial parameters to estimate the x/y shifts for all the tiles in the dataset individually. The beamsplitter and objective lens introduce a consistent distortion pattern that can be approximated with a global affine transform. It is assumed that this is invariant across image tiles. The x/y shifts, however, can vary between tiles as the microscope stage does not always reposition to each tile coordinate in exactly the same spot between rounds, and there can be small errors due to mechanical tolerances or drifts in stage movement gain between rounds.

Hybridisation rounds are registered between channels in a similar manner to sequencing rounds by directly calculating the affine transformation between channels on a tile-by-tile basis using the same block method described above.

Orthogonal matching pursuit (OMP) based gene barcode detection Gene rolonies are detected and matched to the set of known gene barcode sequences by using a modified version of the orthogonal matching pursuit algorithm (Pati et al., 1993) which has been used previously to detect gene barcodes in coppaFISH, an RCA-based in situ sequencing method similar to BARseq (Bugeon et al., 2022).

First, a set of manually selected tiles that contain gene rolony signals are convolved with a difference of Hanning kernel with radii of 2 and 4 pixels (0.462 and 0.924 µm), removing high-frequency noise. Normalisation factors are then calculated per-channel and per-round by computing the distribution of pixel intensities across all channels and rounds and taking the 99.99% quantile. Linear regression in log space is then applied to smooth these normalisation factors and fit the signal drop-off across rounds. Next, the locations of isolated rolony spots are detected from the set of manually selected tiles that contain gene rolony signals. The images are normalised across rounds to the intensity of the first round using the normalisation factors. Then the average fluorescence traces across channels and rounds are extracted from disks with a 2 pixel / 0.462 µm radius in the locations of these spots.

Scaled K-means clustering is used to define each spot as belonging to one of four clusters corresponding to the signal from the four dye-conjugated nucleotides. The cluster means of the four clusters are computed across rounds. By taking a codebook containing the known gene barcode sequences, the mean of these four signal clusters can be used to produce a template of expected fluorescence values in the four channels across rounds for each gene barcode, this is saved as a "gene dictionary".

After assembling this gene dictionary, the OMP algorithm is run on each tile to determine the identity and location of gene spots. For each tile, the difference of hanning kernel is applied, and signal intensity across rounds is normalised to the first round. The gene dictionary and image are then normalised to unit norm. A set of background vectors are created for each channel that each contain fluorescence in just that channel across rounds. The coefficients of the background are calculated and the background is subtracted from the signal to produce the initial residual. Then, the iterative process of OMP is carried out on a per-pixel basis. At each iteration of the algorithm, the component from the gene dictionary fluorescence vectors that has the highest dot product score with the residual of the input signal is selected. This component is kept if the correlation to the residual is more than a defined threshold of 0.2. The coefficients of all the selected components are then refitted via least-squares regression and the residuals are calculated. More gene components can continue to be added in subsequent iterations, enabling OMP to explain the contribution of overlapping gene rolonies to the fluorescence vector of an individual pixel. This iteration continues until a maximum number of components is reached or the correlation to the residual no longer reaches the threshold.

The gene coefficients and residual values for each pixel can then be assembled into images.

Peaks are then detected in these gene coefficient images and a spot score is assigned to each peak based on the similarity of the signal at that location to a precomputed spot image. Any peaks that have a spot score above a certain threshold (0.1), are saved as gene spot locations along with their spot score.

Hybridisation spot detection Hybridisation gene spot detection uses some of the same methods as those used to detect sequencing round gene spots. Peaks are detected in the hybridisation tile images, and then the mean fluorescence values in each channel at a disk with a 2 pixel / 0.462 µm radius in each spot location are extracted. The normalised fluorescence values are used to calculate a dot product with precomputed cluster centroids for that hybridisation round which are mapped to genes by the expected fluorescence of their hybridisation fluorophore probe. The cluster yielding the highest dot product is chosen.

Cell segmentation Cellpose is used to segment cells (Stringer and Pachitariu, 2024). A pre-trained cellpose model (trained by Antonin Blot using DAPI images of mouse visual cortex, based on the "cyto" base model) is used to segment cells, using DAPI images for a nuclei channel and the anchor hybridisation probe that binds to all gene rolonies as a cytoplasmic channel. For low-quality DAPI images with weak/diffuse nuclei signal, cellpose is run in 3D on the non-z-projected DAPI image, and then the resultant masks are projected to a single plane, removing small masks or masks only found on a single plane and merging any remaining.

Image registration between imaging modalities All image modalities are then registered to the first gene round. This aligns the gene spots, hybridisation spots and fluorescence images like DAPI. This is done on a tile-by-tile basis by estimating rotation and translation between the target and reference tiles (using the mean image across channels for both) with phase correlation as described above for registering within channels, across rounds for the gene reference tile.

Tile stitching within ROIs The x/y shifts between each adjacent tile are estimated using the phase correlation of the overlapping regions between the tiles. Large outlier shift values or shifts with low correlation maxima are replaced with the median shift along each tile row. These tile shifts can be used to combine spot coordinates across tiles into a whole ROI coordinate system.

Registration to Allen Atlas Common Coordinate Framework To register the spot coordinates from each ROI into a common coordinate system and determine their location within the brain, all ROIs are registered to the Allen CCFv3 10 µm coronal atlas

(Wang et al., 2020). First, an imaging round of whole brain coronal sections is registered to the ROIs used for sequencing (which are often hemispheres or smaller). This is done by stitching the tiles within each of the two rounds and then finding the x/y shifts between the stitched overview images using masked phase correlation (Padfield, 2012). The whole brain coronal section ROIs are downsampled, and an OME tiff pyramid file is saved for compatibility with the Atlas registration software ABBA (Chiaruttini et al., 2024). All ROIs images are saved into a QuPath project which can be loaded into the Fiji plugin of ABBA. Deepslice (Carey et al., 2023) is used to automate the estimation of atlas cutting angles and positioning of slices along the Z-axis. Elastix (Klein et al., 2010) is then used to perform an affine transformation of the slices. Finally, BigWarp (Bogovic et al., 2016) is used to manually adjust non-rigid transforms to align edges of important structures such as the pia and white matter boundaries of the cortex, the hippocampus and thalamic nuclei. ABBA can output images in the same dimensions as the input ROI images with 3D Allen Atlas coordinates. These are then mapped to the smaller sequencing ROIs through the earlier calculated phase correlation x/y shift. The stitched spot coordinates of each ROI can then be converted to atlas coordinates. The BrainGlobeAtlas API (Claudi et al., 2020) can then be used to find brain area annotations.

Translation to isocortical flatmap depth coordinates The Allen brain atlas coordinates can be converted from 3D CCF data of the mouse isocortex to a flattened cortical representation by using the CCF-streamlines package (Wang et al., 2020). Streamline paths that most directly connect the pia to the white matter of the cortex while following the curvature of the cortex are used to determine the cortical depth of any point within the cortex.

4.3.5 Cell type assignment

Spot assignment to cells Gene spots are assigned to cells by first creating stitched mask images from all tiles in each ROI and creating merged spot coordinates from all tiles in each ROI, using the shifts calculated from registration to the reference to ensure each tile is correctly registered. The whole cell mask image is then dilated by 5 µm to expand the nuclei masks to cover the soma of each neuron. Spots that pass a score threshold of 0.1 and are present within a mask boundary are then added to that mask to create a cell x gene dataframe.

Iterative clustering Cell masks are first assigned to brain regions using their Allen atlas converted centroid coordinates. The following data processing is then performed using the scanpy single-cell analysis toolkit (Wolf et al., 2018). Cortical cells are isolated and filtered to find cells with >=4 unique genes per cell and >=25 gene counts per cell. On these remaining cells, count depth scaling is performed to normalise total counts

per cell to 10, and then the data is logarithmised with a log1p transformation (natural log with a pseudo-count of 1). Principal component analysis (PCA) is then run on the cells using all of the genes. The first 30 principal components are then used to compute the nearest-neighbours distance matrix and a neighbourhood graph of the cells with the local neighbourhood size set to 30. Then louvain clustering is used to cluster cells into subgroups. First, cells are clustered coarsely and the expression of inhibitory marker genes Sst, Vip, Pvalb and Lamp5 along with the excitatory marker gene Slc17a7 are assessed at the cluster level and used to label clusters as excitatory, inhibitory or non-neuronal. Then the excitatory and inhibitory subsets of cells are each split from the other cells and PCA, nearest neighbours and louvain clustering are performed again on only these cells. The neighbourhood graph is embedded in two dimensions with UMAP for visualisation of the final clusters, using a minimum distance between embedded points of 0.5. To the resulting clusters, manual annotation is performed by referencing the expression of marker genes in the gene panel to the V1 SMARTseq reference subclasses (Tasic et al., 2018) and by the location of each cell cluster within the cortex.

4.4 Results

4.4.1 Gene panel selection for maximised classification accuracy

Whilst BARseq provides higher throughput and has lower reagent costs per experiment than techniques such as MERFISH or commercial spatial transcriptomic platforms like Xenium or CosMx, the initial cost of producing a panel of gene probes scales with both the number of genes that are targeted and the number of detection probes selected per gene. This meant that a marker gene panel had to be selected to optimally classify neuronal cell types.

When selecting a limited panel of genes to be used for classifying cells in an in situ sequencing experiment, a number of factors must be considered. First, the detection efficiency of target transcripts can vary from gene to gene but is lower than that of FISH-based approaches (Sun et al., 2021). Therefore, the selection of some traditional cell-type marker genes that are used to delineate scRNAseq-derived clusters may not be applicable as in situ sequencing gene probes due to their low expression values in the populations that they define. Likewise, the selection of too many highly expressed genes in a single in situ panel could lead to optical overcrowding (Alon et al., 2021; Chang et al., 2023; Eng et al., 2019), a phenomenon where the readout capability of the barcode sequences of individual gene transcripts is eventually overwhelmed by the presence of too many transcripts overlapping in space within a cell. Given the typical rolony FWHM diameter of 1 µm (Chapter 3, Figure 3.4 E, H) and a mouse neuronal soma diameter of 10 µm we might expect that when performing data analysis on 2D maximum z-projections, around

90 rolonies fit into a cell's xy plane without overlapping, assuming a hexagonal packing density of $\frac{\pi}{2\sqrt{3}}$:

$$N \approx \left(\frac{10\,\mu\text{m}}{1\,\mu\text{m}}\right)^2 \times \frac{\pi}{2\sqrt{3}} \approx 100 \times 0.9069 \approx 90 \text{ rolonies}$$

In practice, we can use the orthogonal matching pursuit (OMP) algorithm (Pati et al., 1993, Methods 4.3.4) to extract barcode sequence information from overlapping signal sources within an image and this has been used previously to achieve robust readout up to around 250 rolonies per cell (Bugeon et al., 2022). This does place an upper limit on the total expression of gene markers that can be detected within each given cell type, so genes which are highly expressed in many cell types were avoided in my panel. As the aim of this gene panel is to enable cell typing of rabies-infected neurons in the mouse visual cortex and surrounding presynaptic cells, I used a reference SMARTseq dataset which sampled neuronal cell types specifically from V1 (Tasic et al., 2018). Taking into account the minimum and maximum gene expression limitations, the total list of genes from the SMARTseq dataset was filtered to produce a matrix of candidate gene lists that fit a set of minimum expression criteria, plus thresholds on median expression and expression skewness across clusters, with the aim to remove genes with broad cluster expression patterns (Figure 4.1 A). A greedy iterative algorithm was used to select a set of best marker genes from each of these candidate lists. At each step, the algorithm selects the gene which maximises the accuracy of assigning cells from the scRNAseq data to their correct cell type (Methods 4.3.1). From this matrix of gene panels, the panel that maximised the accuracy achievable at a total median geneset cutoff of 30,000 counts was selected (Figure 4.1 B). This cutoff was selected based on the plateau-ing panel accuracy past 30,000 median gene counts and comparison to the total median geneset expression of gene panels that had been previously described at the time of designing this panel from BARseq (Sun et al., 2021) and coppaFISH (Bugeon et al., 2022) experiments. In comparison, when no thresholds are placed on gene expression, a naively selected gene panel using the same greedy algorithm shows only a 1.8% increase in estimated accuracy compared to the thresholded selected panel whilst increasing 49% in median gene expression.

I manually added a set of 16 extra marker genes to the 80 genes selected by the greedy algorithm to include missing canonical marker genes, genes which can be used as alternative markers for cell types with downregulated gene expression of certain markers during rabies infection, and genes which improve the classification of subsets of similar cell types (Methods 4.3.1). Then, I estimated the classification accuracy of the full panel at the "subclass" clustering level of V1 neuronal clusters as defined in the SMARTseq dataset (Figure 4.2 A, B). I also estimated the classification accuracy of the coppaFISH (Figure 4.2 C, D) panel and a 107-gene BARseq panel that has recently been used to classify cell types across the cortex (Chen, Fischer, et al., 2024) (Figure 4.2 E, F) at the same

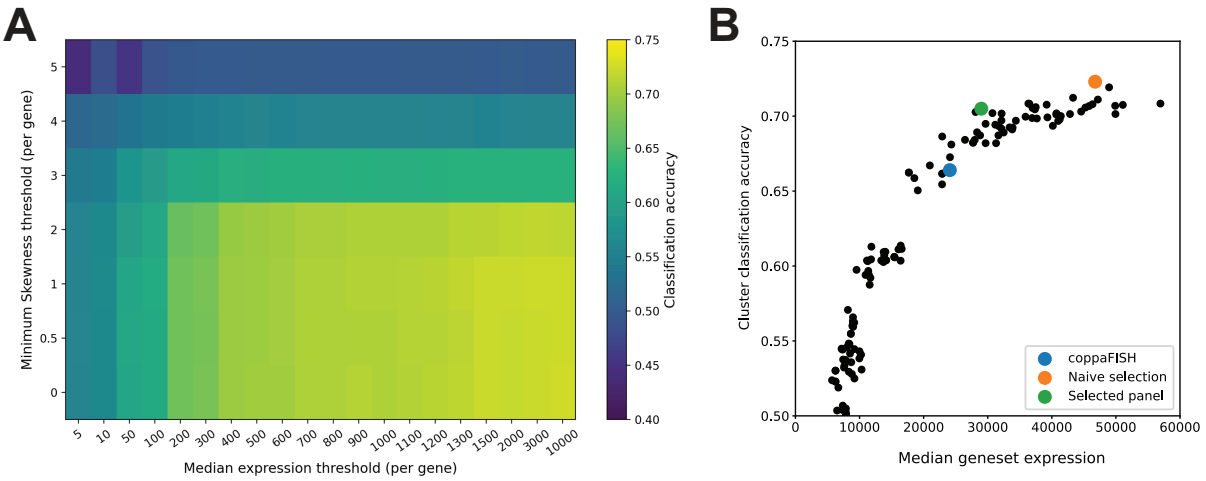


Figure 4.1: Clustering accuracy plateaus with increasing marker gene expression. **(A)** A matrix of estimated classification accuracies for candidate gene panels filtered by a combination of a median expression threshold, a minimum skewness threshold and a fixed minimum mean expression threshold of 500 in the highest expressing cell cluster. Classification accuracy values are the mean accuracy across all cell-type clusters. **(B)** Mean cluster classification accuracy vs total median geneset expression for each of the candidate gene panels. Total median geneset expression is the total of median expression across clusters for all genes in the candidate panel.

"subclass" clustering level. The coppaFISH gene panel is specialised for the classification of inhibitory neuronal cell types in the mouse primary visual cortex and it performs marginally better than my 96 gene panel in classifying these inhibitory neurons (96-gene panel - 96.0%, coppaFISH - 96.4% mean accuracy) but underperforms at classifying excitatory neurons (96-gene panel - 89.2%, coppaFISH - 81.1% mean accuracy). Meanwhile, the 107-gene BARseq panel is designed to classify cells across the mouse isocortex (Chen, Fischer, et al., 2024) and its estimated performance from this SMARTseq dataset of neurons from the visual cortex is lower across all subclasses than my 96-gene panel (96-gene panel - 92.1%, 107-gene BARseq - 76.5% mean accuracy across subclasses).

The SMARTseq dataset is annotated with cell clusters from the 12 "subclasses" that are further refined into 94 "clusters". I estimated the classification accuracy of my 96-gene panel at this "cluster" level. Estimated accuracy varies between individual clusters (Figure 4.3 A). This can mostly be explained due to the misassignment of cells to nearby clusters which are from the same broader subclass (Figure 4.3 B).

4.4.2 BARseq experimental structure

Performing a spatial transcriptomics experiment with BARseq entails first cryosectioning a fresh frozen brain into 20 µm sections which are mounted onto coverslips which can be assembled into flowcells (as described in Chapter 3 and Methods 4.3.3) with ports that allow enzymatic reagents for DNA amplification and sequencing to flow across the samples (Figure 4.4 A). Within these flowcells, the tissue is treated to a series of enzymatic steps

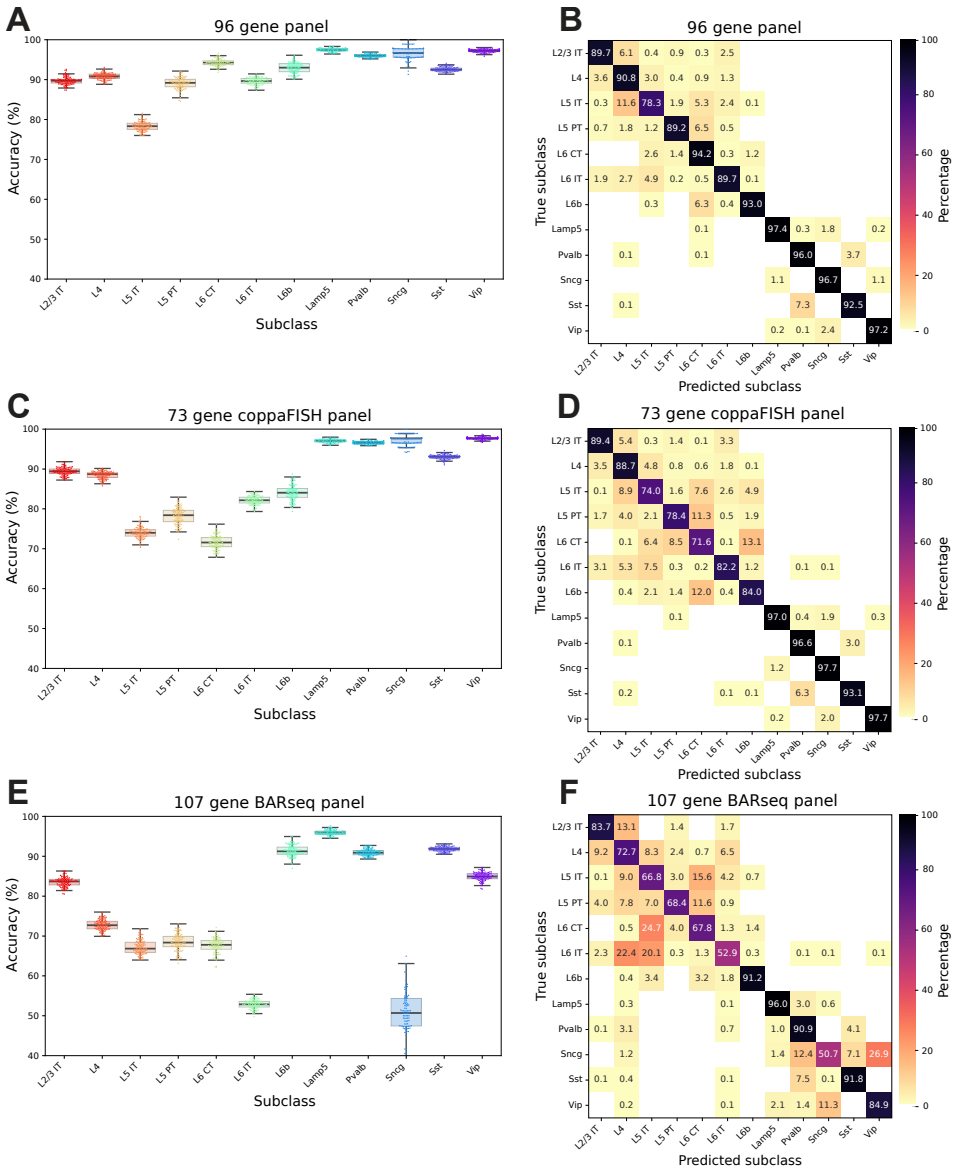


Figure 4.2: The selected gene panel is optimised for discriminating V1 neuronal cell types. (A, C, E) Boxplots of the percentage of cells from each scRNAseq subclass which were correctly assigned to that subclass. Swarmplot points represent each bootstrap permutation of classification accuracy (N=100 bootstraps) and boxes represent the first and third quartiles with median values in bold and whiskers representing the extrema up to 1.5 times the interquartile range of the data. **(A)** shows the accuracy of my 96-gene panel, **(C)** shows the accuracy of the 73-gene coppaFISH panel and **(E)** shows the accuracy of the 107-gene BARseq panel. **(B, D, F)** Confusion matrices of the percentage of cells from each true cell subclass which were assigned to each predicted cell subclass. Values at each true/predicted pair are the median accuracy from all bootstrapped permutations (N=100 bootstraps). **(B)** shows the accuracy of my 96-gene panel, **(D)** shows the accuracy of the 73-gene coppaFISH panel and **(F)** shows the accuracy of the 107-gene BARseq panel.

(Figure 4.4 B, Methods 4.3.3), which first involve adding gene-specific reverse transcription primers and reverse transcriptase to produce cDNA, followed by RNA digestion and the hybridisation of a pool of padlock probes designed to bind to the cDNA of the genes of interest. The adjacent arms of these padlock probes are ligated together to form a circular piece of single-stranded DNA, which is primed for rolling circle amplification into

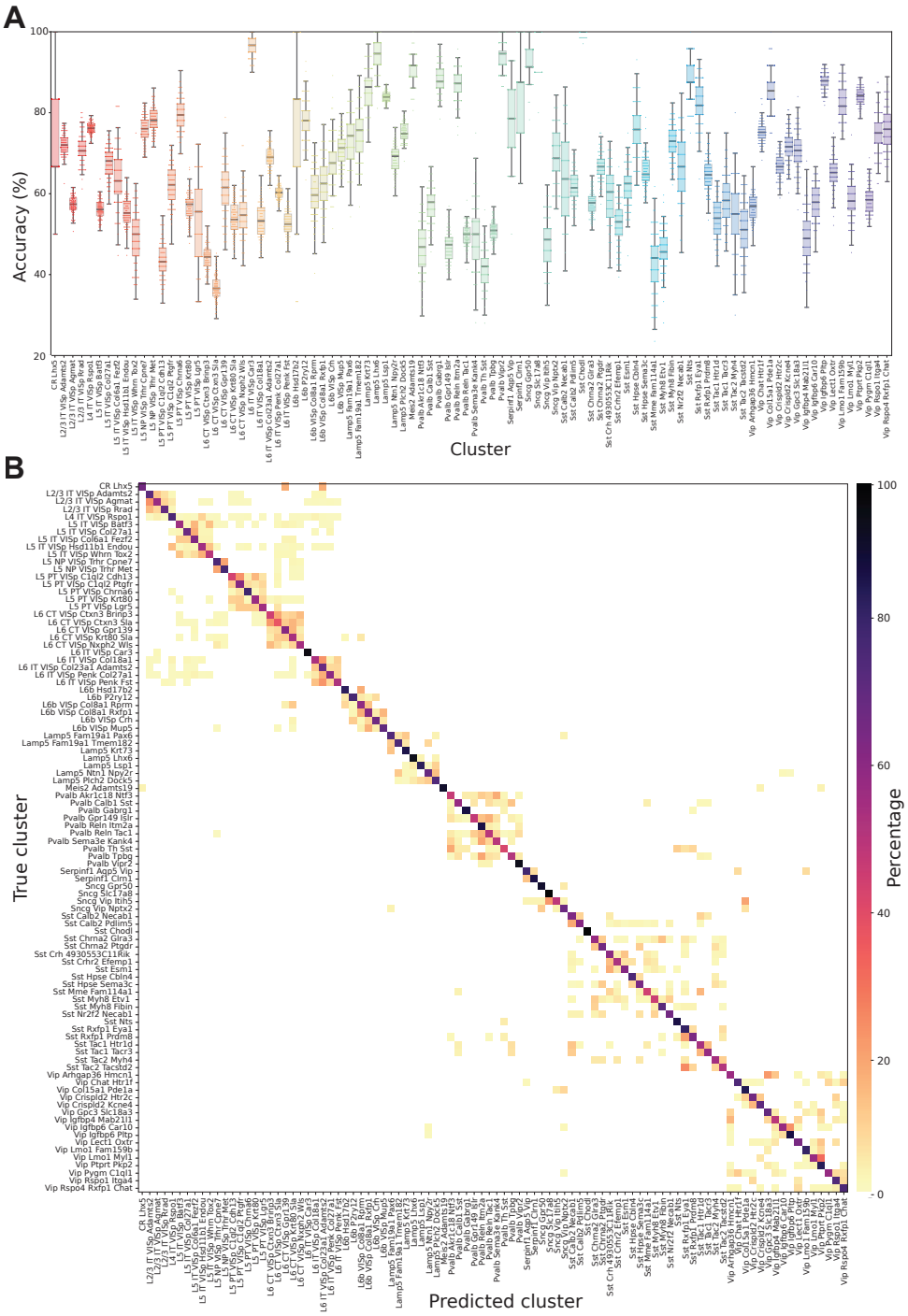


Figure 4.3: Gene panel performance at the cluster level. (A) Boxplots of the percentage of cells from each scRNAseq cluster which were correctly assigned to that cluster. Swarmplot points represent each bootstrap permutation of classification accuracy (N=100 bootstraps) and boxes represent the first and third quartiles with median values in bold and whiskers representing the extrema up to 1.5 times the interquartile range of the data. **(B)** Confusion matrix of the percentage of cells from each true cell cluster which were assigned to each predicted cell cluster. Values at each true/predicted pair are the median accuracy from all bootstrapped permutations (N=100 bootstraps).

a ball of DNA (rolony) containing many copies of the padlock's gene-specific barcode sequence. These rolonies can then be sequenced in rounds of dye-conjugated nucleotide addition, imaging and dye cleavage (as described in Chapter 3 and Methods 4.3.3). The

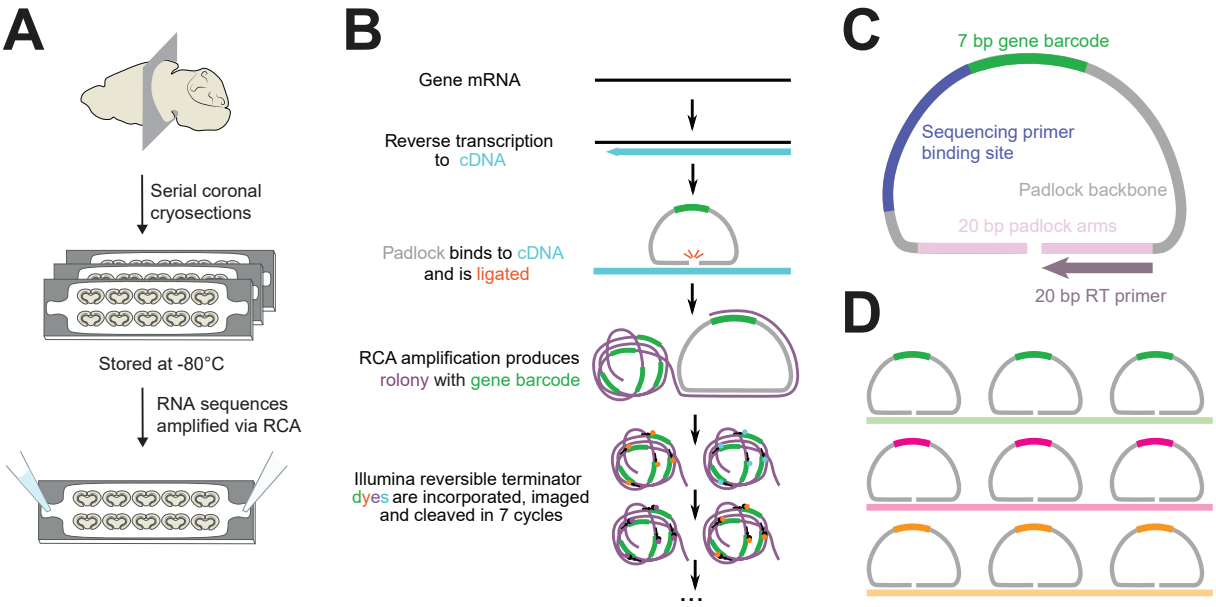


Figure 4.4: In situ sequencing of gene transcripts with barcoded padlock probes. **(A)** A diagram of the in situ sequencing process, starting with a fresh frozen brain which is serially cryosectioned into sections that are mounted onto coverslips, stored at -80C until ready for sequencing when they are assembled into flowcells and RNA sequences are amplified via rolling circle amplification (RCA) **(B)** A diagram of the steps of rolling circle amplification and sequencing. mRNA transcripts are reverse transcribed, then padlock probes with complementary arm regions bind to the cDNA and are ligated. The circularised DNA padlock is then amplified with phi polymerase to produce many copies of the padlock barcode sequence. This is then sequenced using Illumina nucleotide-conjugated dyes in multiple rounds. **(C)** Diagram of the structure of a gene padlock probe. The probe has a 5' and 3' arm region (pink), each of which is 20 bp long and is complementary to the gene of interest. The probe also has a backbone region (grey) which is 40-50 bp long, links the arm regions and contains a sequencing primer binding site (blue) and a 7 bp gene barcode region. A gene-specific RT primer is designed for each padlock which is the reverse complement of the 5' arm region. **(D)** A diagram of padlock probe design, where multiple probes are designed per transcript. All probes that bind to a particular transcript share the same barcode sequence which is specific to that transcript (gene barcode).

padlock probes consist of two arm regions which are designed to bind to adjacent regions of the reverse transcribed cDNA of a gene transcript (Figure 4.4 C), bringing the 5' and 3' ends together. The phosphorylated 5' end enables ligation into a circularised form. Each padlock probe has a conserved backbone region with a sequencing primer binding site and a 7 bp barcode sequence that is unique to each gene. Multiple padlock probes are designed per gene to increase detection efficiency, with each padlock probe for a single gene sharing the same gene barcode sequence (Figure 4.4 D).

4.4.3 Padlock probe design

For each of the genes in the 96-gene panel I designed a set of padlock probes for detecting the genes in situ (Figure 4.5). To increase the detection efficiency of low-expression genes in my panel, I determined the number of probes to be designed for each gene with an inverse relationship to the mean expression level of that gene in its most highly

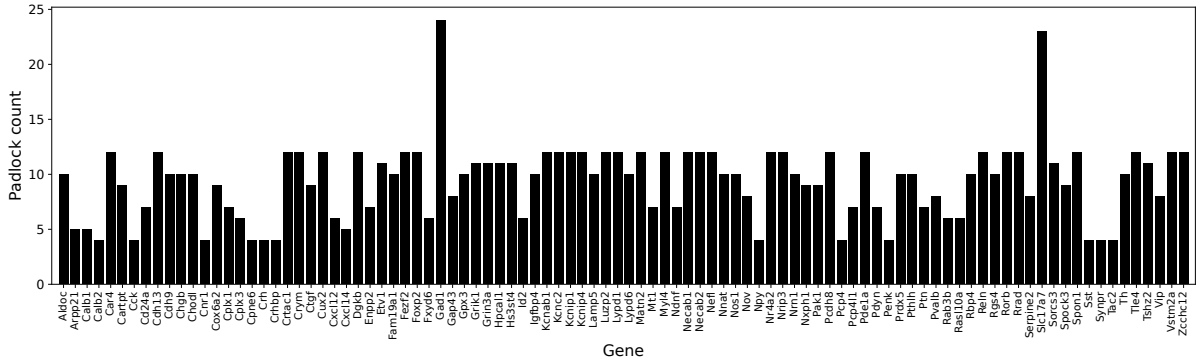


Figure 4.5: Number of padlock probes designed per gene. Barplot showing the number of padlock probes designed for each gene in the 96-gene panel.

expressed subclass (Methods 4.3.1). BARseq and other in situ sequencing techniques that use probes to detect genes do not provide quantitative measures of the differences in expression between the genes that they detect, as each gene will have a different detection sensitivity to every other gene which is based on the accessibility of the probe binding sites for that particular gene. Instead, these techniques provide quantitative information on the differences in gene expression of each of the target genes between cells and regions. Therefore, increasing or decreasing the relative number of probes that target each gene is favourable if it can increase the detection efficiency of weakly expressed genes so that their expression level between cells contributes meaningful information toward classifying cell types. I then selected a set of gene barcode sequences to use for each of the genes in the 96-gene panel. To increase the robustness of detecting the correct gene from barcodes in a noisy experiment, I selected barcode sequences from a previous whitelist of "GII codes" (Sun et al., 2021), which have a minimum hamming distance of 3 between all 96 barcodes (3 base changes away from any other sequence) (Figure 4.6 A). I calculated the minimum hamming distance of the same barcodes if only the first 6 or 5 bases were sequenced and found that all barcodes maintain a minimum hamming distance of 1 even when only 5 bases are sequenced (Figure 4.6 B). For future gene panels that contain larger numbers of genes, I also designed a new set of 1431 x 10 bp barcodes (Figure 4.6 C). These barcodes maintain a minimum hamming distance of 4 at 10 bp sequenced and 2 at 8 bp sequenced (Figure 4.6 D).

4.4.4 In situ detection of gene probes

In order to detect the rolonies generated from these padlocks after RCA (Figure 4.4 B, Methods 4.3.3), a series of sequencing rounds are performed, where dye-conjugated nucleotides are extended from a sequencing primer across the gene barcode sequence, reading the bases one by one as fluorescence in one of the 4 imaging channels of the BARseq microscope (Figure 4.7 A). The following image processing steps are discussed in detail in Methods 4.3.4. Isolated rolony spots are found from an image and scaled k-means

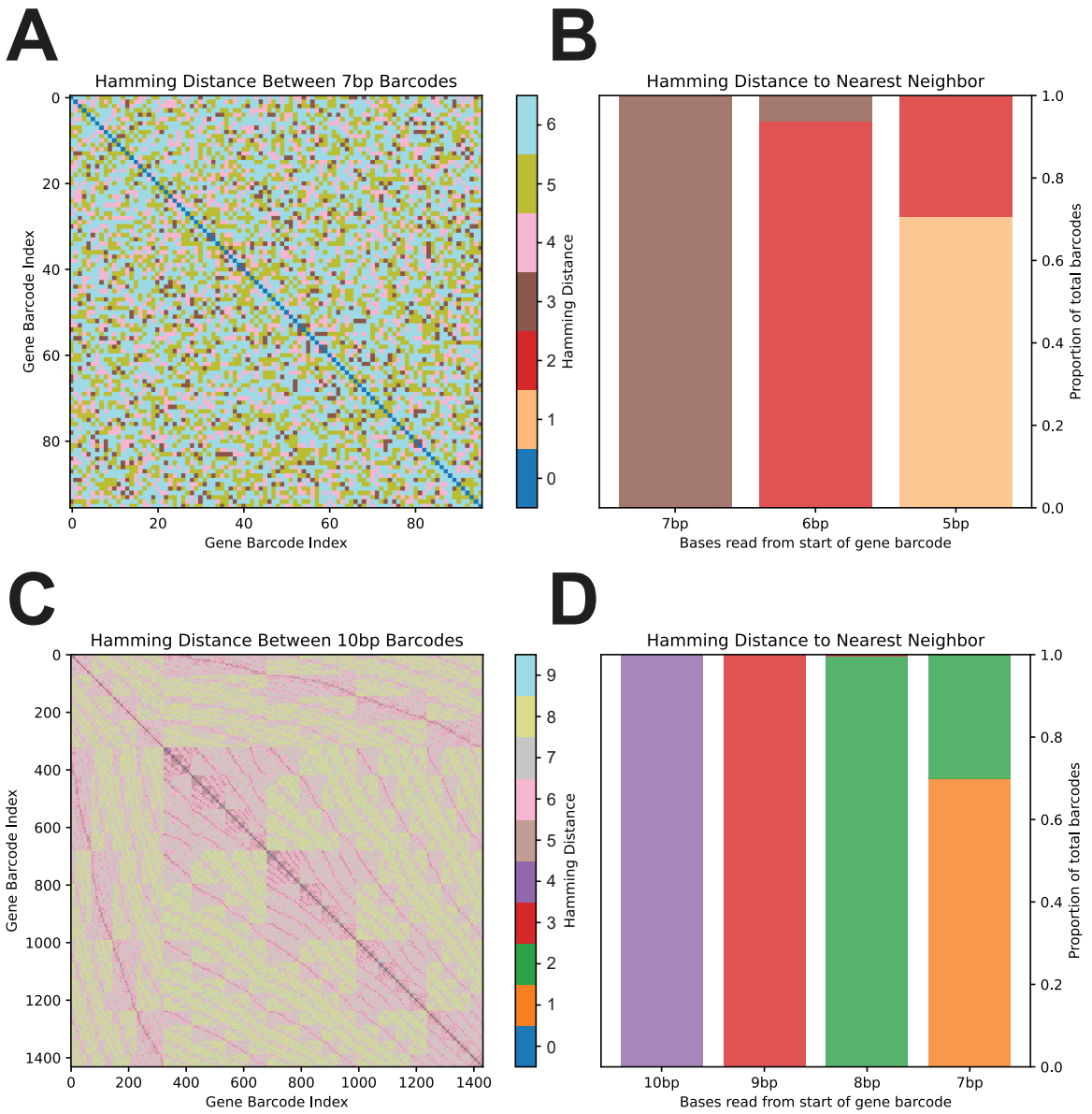


Figure 4.6: Gene barcode sets are optimised for maximum hamming distance. (A) Matrix of the pairwise hamming distance between each of the 96 gene barcodes used in the 96 gene panel. **(B)** Stacked bar plots of the hamming distance to the nearest neighbour barcode for the 96 gene barcodes. Each plot shows the proportion of nearest neighbour hamming distances when reading the whole 7 bp barcode or a subset of the first X bases. **(C)** Matrix of the pairwise hamming distance between each of the 1431 gene barcodes generated for an optimised 10 bp barcode list. **(D)** Stacked bar plots of the hamming distance to the nearest neighbour barcode for all 1431 of the optimised 10 bp barcode list. Each plot shows the proportion of nearest neighbour hamming distances when reading the whole 10 bp barcode or a subset of the first X bases.

clustering (Bugeon et al., 2022) is performed on the fluorescence values from each spot to define the cluster means from the 4 possible dye-conjugated nucleotide signals. This is performed across rounds to produce an observed bleedthrough matrix of fluorescent signal for each base (Figure 4.7 B). This is then used to assemble a dictionary of expected fluorescence traces across rounds and channels for each of the gene barcodes in the panel (Figure 4.7 C). Using the OMP algorithm (Bugeon et al., 2022; Pati et al., 1993), the

contribution of these expected fluorescence traces to the observed signal found at every pixel in the microscope images is estimated. This enables the location of gene spots to be detected from peaks in the resultant pixel coefficients for each gene. DAPI was detected in a further round of imaging and this nuclear stain was used to define cell masks with the deep learning-based cell segmentation tool - Cellpose (Stringer et al., 2021). I then dilated the resultant cell masks and assigned gene spots within the boundary of these masks to those cells (Figure 4.7 D). The use of DAPI to segment cell boundaries is not ideal for assigning transcripts to cells. Whilst DAPI is useful for providing a reference location of the centre of a cell, neuronal gene transcripts are spread throughout the cell body and in the dense tissue of the brain, so masks must be expanded to prevent the majority of transcripts from being unassigned to cells. However, nearby expanded DAPI masks can incorporate rolonies from neighbouring cells into each other's cell masks, which hinders subsequent clustering analysis. In Chapter 5, I explore this further with alternative methods for segmentation which could be used to generate cell masks that delineate neuronal cell bodies (Figure 5.10 E, F, Discussion 5.5).

To qualitatively confirm the specificity of the individual gene probes, I compared the coordinates of detected gene rolonies from an experiment that covered whole mouse brain coronal sections at Bregma -2.9 mm (a region that includes anterior primary visual cortex and higher visual areas) to in situ hybridisation stains of matched coronal sections from the Allen ISH atlas (Lein et al., 2007) (Figure 4.8). Spatial patterns of the genes were consistent with the Allen ISH staining.

4.4.5 Identifying transcriptomic cell types in situ

The following analysis is from 40 serial coronal 20 µm cryosections that I collected, spanning 800 µm of tissue that is centred around the injection site of a series of mediolateral barcoded rabies virus injections into the mouse primary visual cortex and surrounding higher visual areas. The imaged region of each of these coronal cryosections is about two-thirds of a hemisphere each. 1,954,128 total cell masks were detected across all of these cryosections. Of these cell masks, 588,568 cell masks had less than five gene rolonies assigned to them. Some of this is likely due to spurious mask creation from the DAPI signal, whilst another large majority of these cells are from midbrain and non-cortical structures where expression of my cortical marker gene panel is low. Another population of cells with few to no transcripts are non-neuronal cells including astrocytes, oligodendrocytes, microglia and endothelial cells, for which this panel contains no specifically targeted marker genes. For the purposes of classifying cell types via de novo clustering, I used cell masks located only in the isocortex and set an inclusion cutoff at 25 gene counts per cell and 5 unique genes per cell for clustering analysis (Methods 4.3.5). This resulted in the selection of 969,905 cells, including all cells from isocortex. After count filtering, 206,756 cells were left.

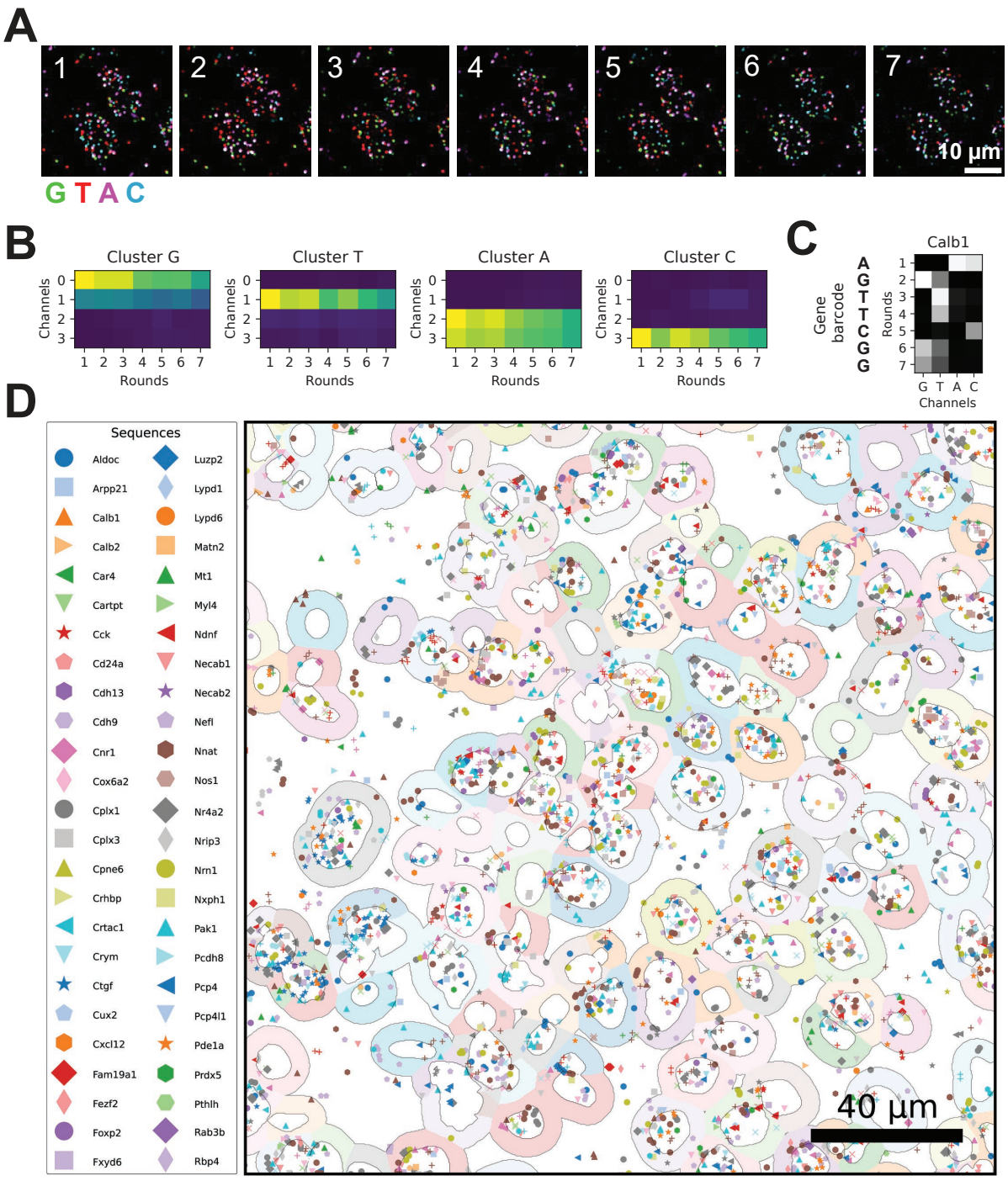


Figure 4.7: Gene detection and cell assignment pipeline. **(A)** Representative four-channel microscope images of the fluorescent nucleotide dyes used for sequencing the gene barcode across seven imaging rounds. Images have been aligned between channels and convolved with a Hanning kernel (scale bar - 10 µm). **(B)** Bleedthrough matrices showing the mean rolony intensity values across channels and rounds for each of the four k-means separated clusters of nucleotide dyes. Values for each channel are normalised to the intensity of the first round of imaging. **(C)** An estimated fluorescence matrix for the Calb1 gene barcode, determined from the observed cluster mean intensities of **(B)** for each round/channel. **(D)** Detected gene coordinates (symbols) plotted over DAPI masks (white centre) expanded by 5 µm (random coloured fill per mask). All gene spots within the expanded masks are assigned to that cell mask for cell-type clustering.

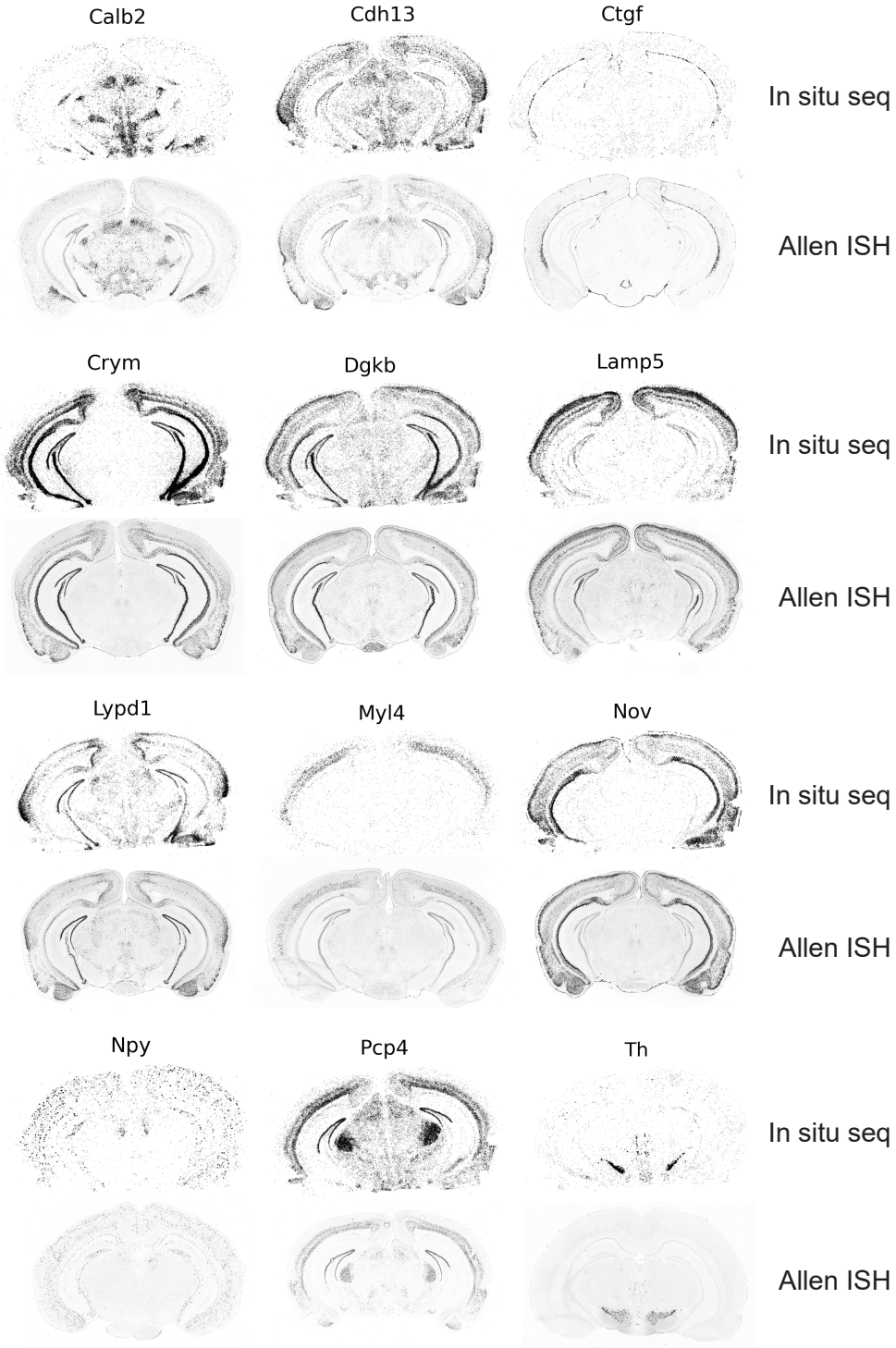


Figure 4.8: Spatial patterns of genes detected in situ match existing ISH data. A comparison of a selection of genes detected with in situ sequencing and the same region of the brain from coronal ISH data derived from the Allen ISH atlas.

I proceeded to iteratively cluster the cortical cells, first into coarse clusters of excitatory, inhibitory and non-neuronal cell types, before reclustering the excitatory and inhibitory cells separately (Methods 4.3.5). This iterative re-clustering process enables expression variability which is specific to excitatory or inhibitory cells to be represented during the dimensionality reduction and clustering process, increasing the resolution of distinguishable clusters (Luecken and Theis, 2019; Nehar-Belaid et al., 2020; Zheng et al., 2022). I

then manually annotated these final clusters (Figure 4.9 A), naming them according to the corresponding subclass conventions from the mouse V1 SMARTseq reference data (Tasic et al., 2018) by using a combination of the observed expression patterns of each cluster compared to the SMARTseq data and spatial patterns of the cells in the cortex. All major inhibitory cell subclasses were detected apart from Sncg, and all excitatory cell subclasses were detected apart from L5 NP and a separable L5 IT (only L4 and the broader L4/5 IT class could be separated). I explored common covariates which may affect clustering, such as total counts per cell (Figure 4.9 C) and unique genes per cell (Figure 4.9 D). None of the defined clusters shows consistently high or low total transcript counts or gene counts that would be expected from technical artefacts driving the clustering results. Instead, each cluster demonstrates a gradient of transcripts and genes found per cell, indicative of high and low-quality populations of detected cells in each cluster. These could be due to variations in the quality of cell masking and subsequent spot assignment, or they could also be due to cells which are varying in depth within each cryosection and have varying proportions of the cell cut out of plane, resulting in variations in detected spot counts. Finally, it could be due to variations in detection efficiency between cryosections or flowcells. Each of the four flowcells in this experiment does not show strong clustering patterns in UMAP space (Figure 4.9 B). When comparing the total counts per cell (Figure 4.9 E) and unique genes per cell (Figure 4.9 F) between flowcells, no major differences are detected, and the same applies for the proportion of detected cell type clusters (Figure 4.9 G).

When comparing the expression patterns of my gene panel across the subclasses in the reference SMARTseq data (Figure 4.10 A) to the matching clusters in my in situ sequencing dataset (Figure 4.10 B), a number of genes that have expected expression across some clusters are not detected at high levels in any of the cell clusters in the in situ sequencing data (*Cartpt*, *Crh*, *Cxcl12*, *Gpx3*, *Matn2*, *Penk*). These likely represent genes with probes that have a low detection efficiency for those transcripts. A larger proportion of the genes do demonstrate expression patterns across the subclass clusters which match those of the reference dataset closely. *Ctgf*, *Npy*, *Pvalb*, *Sst* and *Vip* are all exclusively found in single subclasses, as in the reference dataset. *Chgb*, *Cplx1*, *Lamp5*, *Pcp4*, *Pde1a*, *Rgs4* and *Slc17a7* likewise match the expression patterns across multiple subclasses. Other genes such as *Calb1*, *Cck*, *Cplx3*, *Gap43* and *Nov* exhibit weaker expression pattern matches to the reference data, which could be attributed to low detection efficiency or to discrepancies in the clustering of the in situ data compared to the reference data.

Another confounding factor in the data from in situ sequencing is the possibility of reads from two cells that are in close proximity being misassigned to the mask of a single cell, if these adjacent cell masks are fused, analogous to cell doublets that are found in scRNAseq data. In another misassignment scenario, it is also possible that transcripts that are detected in passing neuronal processes from the neuropil could be incorrectly assigned to the masks of random cells. To measure the rate of these occurrences, I compared the expression of the *Vip*, *Sst* and Parvalbumin (*Pvalb*) genes in the expected Vip, Sst and

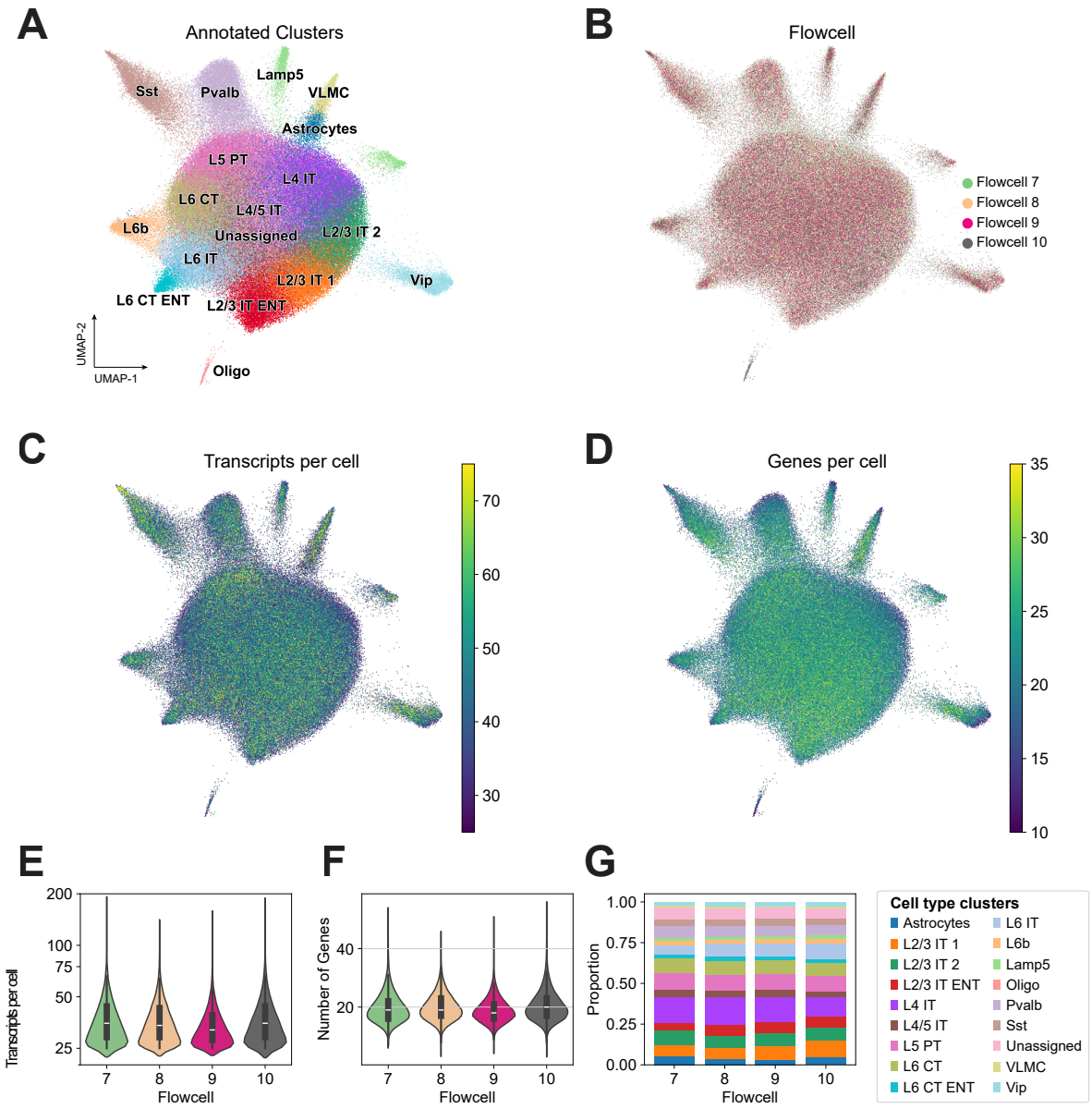


Figure 4.9: De novo cell clustering from in situ detected transcripts is consistent across flowcells. UMAP representation of cells, coloured by assigned cell type cluster **(A)**, flowcell of origin **(B)**, transcripts detected per cell **(C)** and unique genes detected per cell **(D)**. **(E)** Violin plots of transcripts per cell for each flowcell. **(F)** Violin plots of genes per cell for each flowcell. **(G)** Stacked barplots of the proportion of cell type clusters found in each flowcell. VLMC: vascular leptomeningeal cells.

Pvalb inhibitory cell types that they should be exclusively expressed in, to their expression in all glutamatergic cells. In the reference scRNAseq data, expression of inhibitory marker genes is high within their respective clusters and low in the glutamatergic cells (Figure 4.11 A). In my in situ sequencing data, the expression is also high within the expected clusters, but 12-21% of the glutamatergic cells also express these inhibitory cell markers at low count levels of 1-2 counts per cell. These low-level count values could be from individual rolonies from adjacent cells or from neuronal processes. In the case of Parvalbumin this could also be genuine low-level expression which is found in some L2/3 glutamatergic

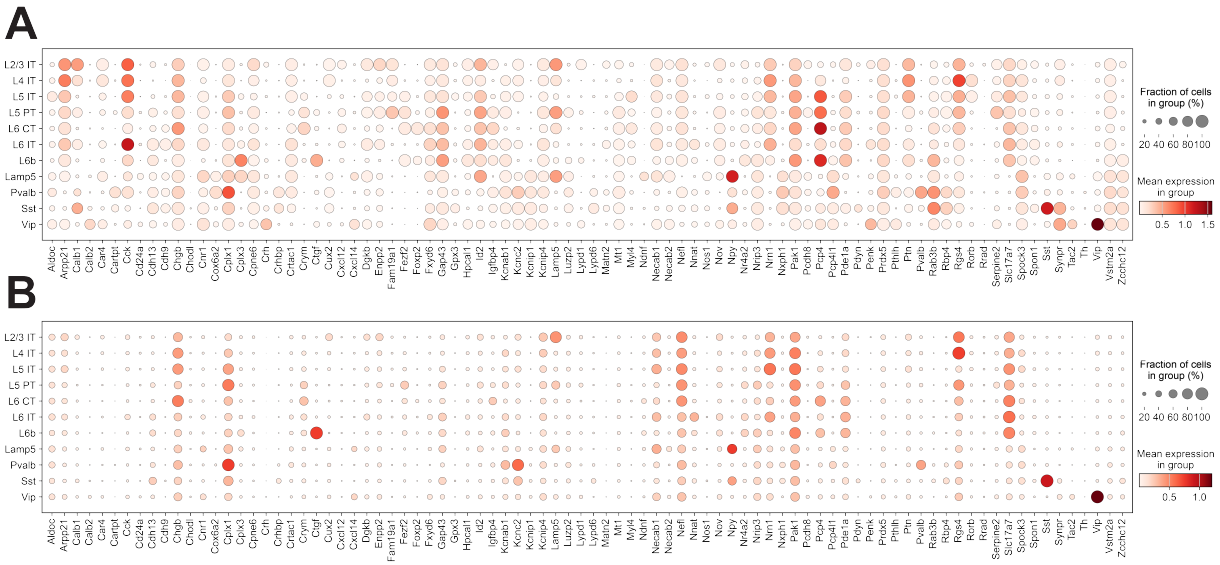


Figure 4.10: Gene expression patterns of de novo cell clusters from in situ data follow scRNAseq cluster expression. (A) Dotplot of the SMARTseq expression values of the marker genes in my 96-gene padlock probe set across scRNAseq-defined cell type subclasses from mouse V1 (Tasic et al., 2018). **(B)** Dotplot of the in situ sequencing expression values of my 96-gene padlock probe set across manually annotated cell type clusters that match the scRNAseq subclass clusters of **(A)**. Circle size represents the percentage of cells expressing each gene per cluster and circle colour represents the expression level of the gene. Expression values are count depth normalised and log1p adjusted.

populations (van Brederode et al., 1991).

The fidelity of the resultant de novo cell clusters can also be assessed spatially by comparing them to known laminar distributions, especially for cortical excitatory neuronal subclasses. Each of the defined subclass clusters follows the general laminar patterns that are expected of them, with L2/3 IT types present near the cortical surface, and deeper L4 IT cells succeeded by L4/5 IT, L5 PT, L6 IT/CT and L6b cells (Figure 4.12). However, quantifying the distribution of these cells is difficult when working in the curved structure of the cortex. For this reason, I transformed the coordinates of the Allen CCF-mapped cells in the cortex to an isocortex flatmap coordinate space so that linear cortical depth information could be extracted and compared for individual cells (Methods 4.3.4).

Transforming the coordinates into a flatmap also enables a 2D visualisation of the physical extent of the in situ dataset within the cortex (Figure 4.13). The range of cortical areas that the dataset covers is apparent from this map. Cells are found in the primary visual cortex, the higher visual areas VISpm, VISl, VISli, retrosplenial cortex, auditory cortex, temporal association area and beyond.

I selected a subregion of cells within the bounds of the primary visual cortex for analysis of the cell type distribution across cortical depth from the flatmap depth coordinates (Figure 4.14 A). This was to avoid changes in the thickness of the cortex itself and changes in the composition of cell types that occur in more medial/lateral areas. As seen in (Figure 4.14), the distribution of glutamatergic cell types follows the expected laminar pattern of

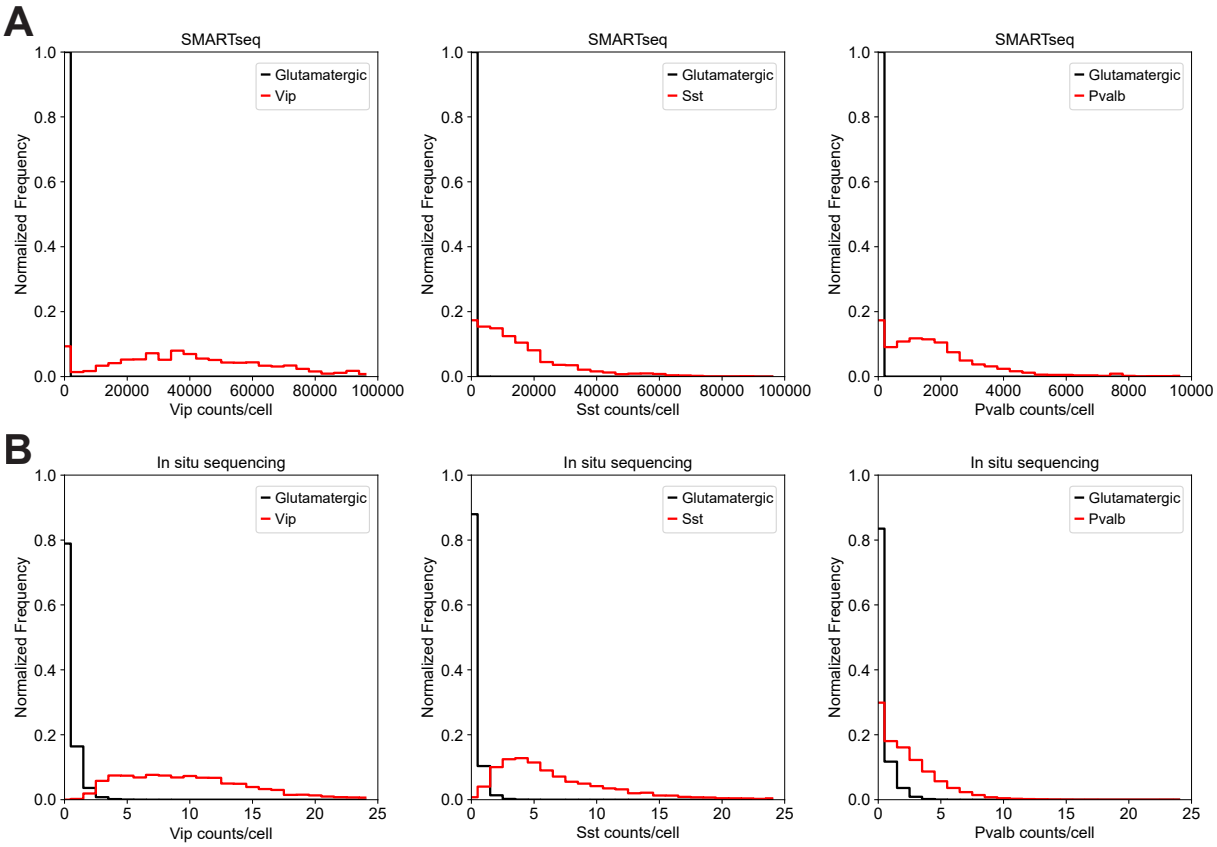


Figure 4.11: In situ misassignment of GABAergic marker gene transcripts to glutamatergic cells. (A) Distribution of SMARTseq exon read counts of GABAergic marker genes Vip, Sst and Parvalbumin in their respective subclass clusters (red) and in glutamatergic cells (black) in scRNAseq reference data (Tasic et al., 2018). **(B)** Distribution of in situ expression values of GABAergic marker genes Vip, Sst and Parvalbumin in their respective de novo in situ cell type clusters (red) and in glutamatergic cells (black). Glutamatergic cells in both plots are defined as the union of L2/3 IT, L4 IT, L5 IT, L5PT, L6 IT, L6 CT and L6b cells.

the layers in the visual cortex (Figure 4.14 B). The GABAergic and non-neuronal cells also followed expected layer patterns. The VLMC (vascular leptomeningeal) cell cluster is present almost exclusively at the pial surface as expected (Pietilä et al., 2023), and the expected bias towards superficial layers for Lamp5 and Vip interneurons compared to a deeper layer bias for Pvalb and Sst interneurons (Gouwens et al., 2020; Scala et al., 2021) is also visible (Figure 4.14 C). However, these spatial distribution patterns reveal a substantial number of cells that have been assigned to subclasses despite being far from their expected laminar location. To determine the proportion of cells which spatially matched the expected layer of their assigned transcriptomic cell type, the mean cortical depths of the layer boundaries of the visual cortex (Wang et al., 2020), were used to define the extent of each layer. Cells were classified as within their expected layer if they fell between the correct layer boundaries, dilated by +/- 100 µm for each layer to account for mouse-to-mouse variability in layer thickness. Whilst 91% of the excitatory cells assigned to the L2/3 IT clusters are found within the correct layer boundaries, only 60% of the cells assigned to the L4 IT cluster are present in that layer (Figure 4.14 D).

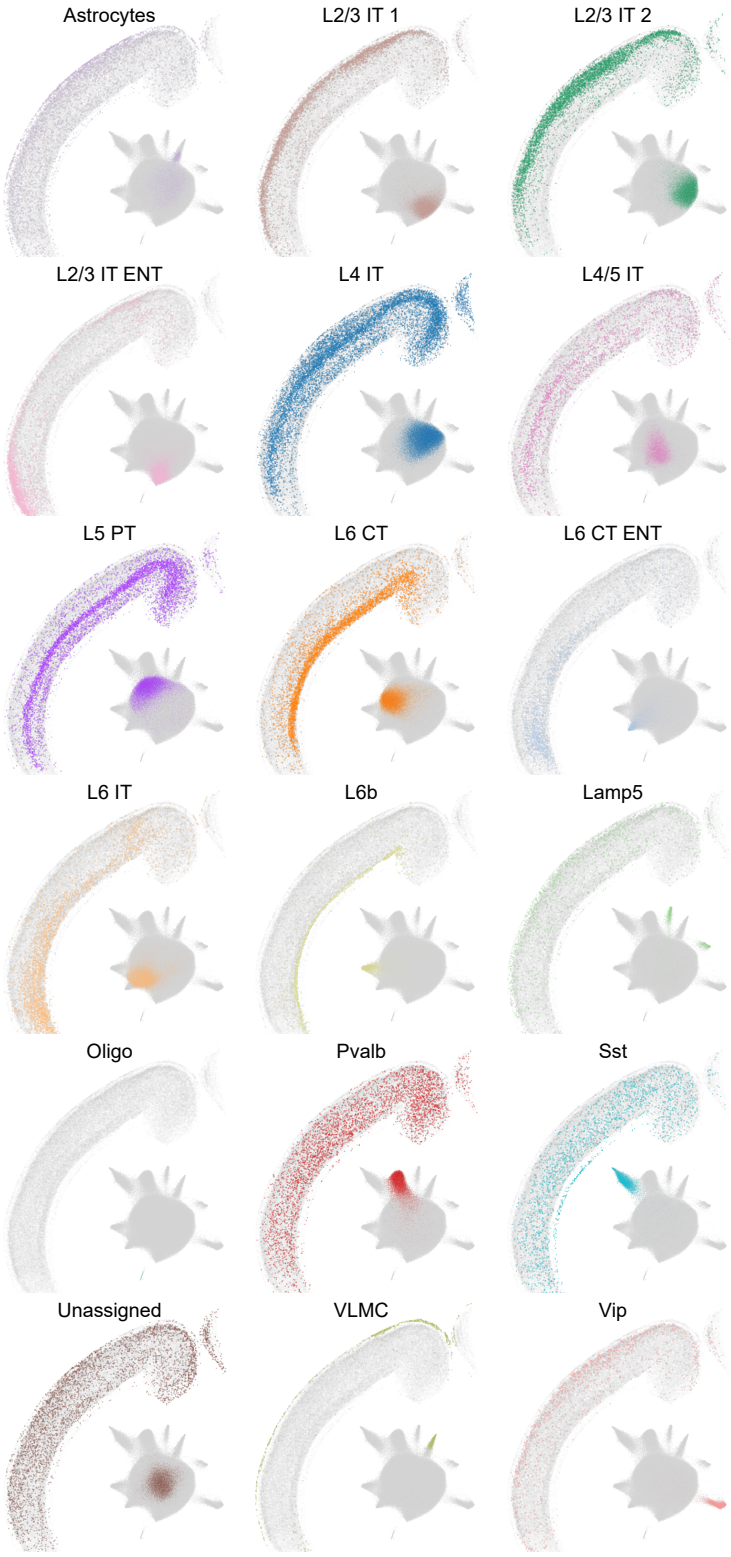


Figure 4.12: Spatial distribution of de novo cortical cell types from in situ iterative clustering. Spatial plots of cell locations of each cell type cluster from in situ sequencing with inset UMAP plots highlighting each cluster. Cell locations are derived from a representative stack of 10 serial brain sections projected through the anterior-posterior axis.

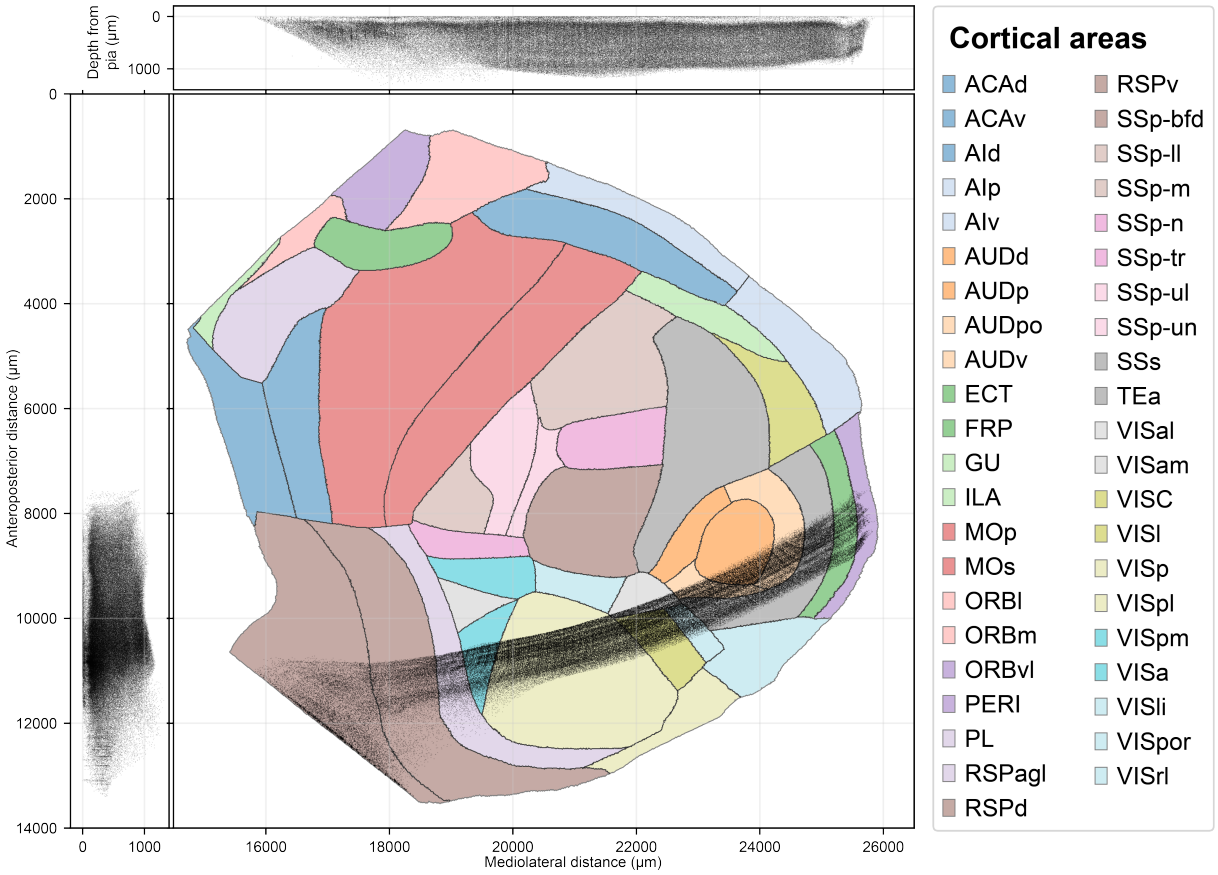


Figure 4.13: In situ sequencing data can be mapped to a common coordinate framework. A dorsal flatmap view of the Allen CCFv3 brain atlas with cortical neuron coordinates from the in situ sequencing data mapped into flatmap coordinate space with laminar depth from the pia to white matter calculated from isocortical streamlines (Wang et al., 2020). The central dorsal view denotes Allen cortical areas in different colours with cell coordinates in black with a cross-sectional upper mediolateral projection and a left anteroposterior projection.

4.5 Discussion

I have shown in this chapter that it is possible to design a gene panel with a limited number of genes which still achieves a high estimated accuracy across almost all the transcriptomic clusters of neurons in the visual cortex. I have demonstrated that the genes in this panel can be detected in situ and that probe sets for individual genes demonstrate the specific expression patterns that are expected. I have shown that these detected genes can be assigned to cells and that even with non-ideal segmentation and detection efficiencies, the repeatability of clustering across flowcells is consistent and the patterns of expression within de novo clustered cells largely match expected patterns from reference sequencing data. These de novo clusters also largely exhibit the spatial distributions expected of each cell type. By improving the remaining issues of segmentation, increasing gene detection efficiency and performing more quality control analysis on the de novo clusters, the accuracy of transcriptomic cell type assignment could be further improved for future experiments.

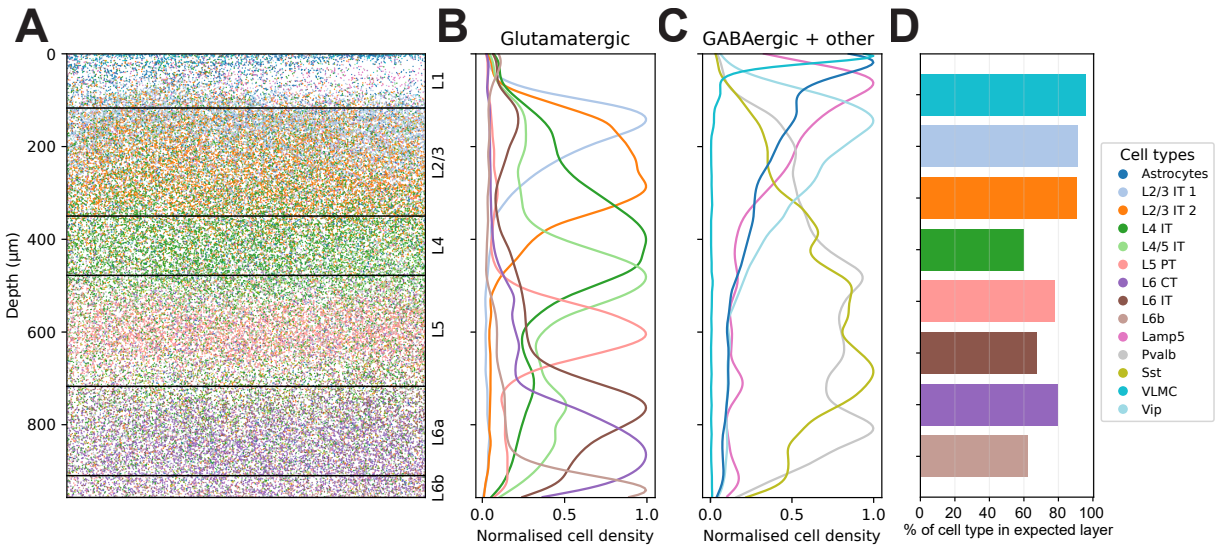


Figure 4.14: The laminar distribution of cell types in situ predominantly follows expected layer patterns. (A) Spatial distribution of cortical cells mapped into a flattened coordinate system with normalised cell depth from the pia to white matter boundaries. Black horizontal lines denote the average depths of layers calculated from measurements in the mouse visual cortex from Allen atlas CCFv3 (Wang et al., 2020). **(B, C)** Distribution of cell depths for each glutamatergic cell type **(B)** and each GABAergic or non-neuronal cell type **(C)**. Each line represents a kernel density estimation of the cell depths for a cell type with density peaks normalised to 1. **(D)** Percentage of each cell type that is found within its expected layer (+/- 100 µm).

There are several factors that can decrease the percentage of cells which are accurately assigned to their correct cell type during an in situ sequencing experiment. Regarding gene panel design, we must use a limited amount of information in the form of a targeted gene panel to determine these cell types in situ. Before considering any experimental challenges, my prior estimates of classification accuracy based on the information available from downsampling read counts from scRNAseq reference data for a selected set of 96 genes, gives the classification of some distinct cell types at close to perfect accuracy (Lamp5 - 97.4%, Vip - 97.2%), meanwhile, other cell types already begin with a comparatively low classification accuracy (L5 IT - 78.3%, with 11.6% being misclassified as L4 IT cells).

Another factor which affects transcriptomic classification accuracy is the efficient detection of the targeted genes. A common strategy with in situ sequencing techniques is to design multiple gene probes per gene to increase detection efficiency and avoid the failure which may occur to a few individual probes. I observed that some genes in my gene panel were detected with low efficiency. The probes for a large number of the other genes in my panel produced highly specific spatial distributions of detected signal throughout the brain which follow the patterns that have been detected previously with ISH. The genes for which detection failed may be because I designed small numbers of probes for some of the more highly expressed genes (as low as 4 per gene), or because those genes are harder to detect due to problems of RNA secondary structure, degradation of transcripts, or failure of reverse transcription for some templates. While the gene panel was designed

to avoid optical overcrowding from selecting too many highly expressed genes, with the gene detection efficiency achieved in these experiments, this has been largely unnecessary, with only the most highly expressed genes overcrowding the cell (Sst, Vip), and these were imaged in a separate hybridisation round to prevent this from affecting gene calling. This means that at these detection efficiencies, a sufficient margin remains for adding more highly expressed genes to the panel to increase accuracy of classifying cell types.

BARseq and its precursors, BaristaSeq (Chen et al., 2018) and ISS (Ke et al., 2013) are techniques which rely upon reverse transcription to detect transcripts. The padlock probes of all three approaches are designed to bind to the cDNA sequences of reverse-transcribed transcripts, which enables the use of DNA/DNA ligases to circularise the padlock probes. More recently, alternative rolling circle amplification (RCA) based approaches have managed to circumvent the need for reverse transcription and have demonstrated increased detection efficiency of transcripts as a result (Lee et al., 2022; Sountoulidis et al., 2020). These techniques include SCRINSHOT (Sountoulidis et al., 2020), BOLORAMIS (Liu et al., 2021), targeted ExSeq (Alon et al., 2021) and dRNA-HybISS (Lee et al., 2022) which all use the Chlorella virus DNA/RNA ligase (SplintR) to enable ligation of padlock probes that are directly targeted to RNA transcripts. Another method, STARmap (Wang, Allen, et al., 2018) uses SNAIL probes which are comprised of a padlock and secondary DNA oligonucleotide that bind in tandem to the RNA transcript, with the secondary oligonucleotide bridging the two padlock arms and avoiding the need for DNA/RNA ligation. Whilst the increased detection efficiency of these approaches makes them appealing for use in detecting the sequences of known gene transcripts, my experiments also require the determination of sequences of unknown rabies barcodes. Compared to detecting know gene sequences, this requires a step of gap-filling the unknown rabies barcode sequence into the barcode padlock so it can be amplified and sequenced. A version of these reverse transcription-free techniques that enables RNA binding padlock probes to gap-fill arbitrary sequences like unknown viral barcodes has yet to be described. It is possible that such a technique could be enabled by performing gap-filling from the RNA template, priming from the DNA padlock arm with a reverse transcriptase. If a reverse transcriptase with reduced RNase activity is used, like Superscript IV, the SplintR ligase could act upon the resultant RNA splint to circularise the padlock. Given the reduced fidelity of the SplintR ligase for single base mismatches (Krzywkowski and Nilsson, 2017) and its bias for certain donor/acceptor bases at the ligation site (Lohman et al., 2014; Schneider and Meier, 2017) along with the lack of pre-established RNA gap filling methods that were available at the time of conception of my own experiments, I designed my gene panel around the already proven BARseq technique. Using my proposed RNA gap-filling method described above or by using gapfilling BARseq for rabies barcode detection in complement with one of the established direct RNA padlock approaches for non-gap-filling gene detection, it is possible that transcripts that are not reliably detectable with BARseq could be detected in future experiments.

Another cause of reduced data quality that affects the resultant classification accuracy of cells comes from cell segmentation. Segmentation is crucial for defining the identity of each cell, as inaccurate segmentation can result in: over-splitting masks into multiple false cells; under-splitting cells - which result in cell doublets; producing undersized masks - resulting in missing rolonies from cells; or overextended masks - resulting in incorporating rolonies unrelated to a cell. All of these compromise the accuracy of cell expression patterns. Currently, the use of DAPI as a segmentation signal is limiting as it only captures the nucleus of the cell, whilst many of the transcripts in neurons reside in the soma which can take on elongated/pyramidal shapes. Using the anchor signal of all gene rolonies as a substitute cytoplasmic stain may help somewhat, but at the current detection efficiencies that I achieve with BARseq, this signal is too sparse to reliably be used to delineate cell boundaries. I have since implemented antibody staining post-sequencing (Figure 5.10 E, F), and the use of cell body/membrane antibody markers should provide a useful tool for improving the segmentation of cortical neurons significantly in future experiments.

Finally, further data processing steps could be used to improve the percentage of cells which are correctly assigned to a cluster. After clustering has been performed, it would be possible to generate clustering quality control scores for each cell to determine the confidence that it has been assigned to the correct cluster. Currently, a cell must only pass thresholds for a minimum gene count and a minimum number of detected transcripts to be included in the iterative clustering approach. In the future, cell doublets and low-quality cluster assignments may be removed from further analysis by determining their correlation to the gene expression values of each cluster centroid or other quality control metrics. Direct mapping from scRNAseq clusters to spatial transcriptomic data has been performed on other datasets (Yao, Wang, et al., 2023) but the intrinsic technique-specific differences in the readout of transcripts can lead to artefacts when integrating data crossmodally (Liu et al., 2023). Given a fixed gene panel, a selection of high-quality cell clusters from one experiment could be used to perform cluster mapping from a reference dataset to future datasets.

Chapter 5

Determining connectivity with barcoded rabies viruses

5.1 Authorship declaration

I, Alexander Becalick, performed the majority of the experiments and analyses in this chapter with the exceptions indicated here. Antonin Blot and Sophie Wood of the Francis Crick Institute performed the surgical procedures on the mice. Petr Znamenskiy and Antonin Blot developed the data preprocessing pipeline to base call rabies barcode sequences and assign them to cells. Antonin Blot performed imaging and automated cell counting for the serial sectioning two-photon data.

5.2 Introduction

Resolving the connectivity of neurons using a barcoded rabies labelling strategy requires several experimental conditions to be fulfilled in order to produce a robust representation of the underlying connectivity. In Chapter 2, I covered the production of libraries of barcoded rabies viruses with diverse barcode sequences. From quantifying the barcode abundance distributions of these libraries, I have found that they are theoretically capable of uniquely labelling thousands of initially infected starter cells in a single experiment. To determine the synaptic inputs onto the large populations of neurons that could be labelled with such libraries, it is necessary to detect and sequence rabies barcodes from infected cells. This can be achieved in a similar manner to the in situ detection of endogenous transcripts that I describe in Chapter 4. These barcodes must be correctly assigned to their cells of origin using the spatial information that is preserved with this in situ sequencing method. The direction of synaptic connectivity between these cells can then be determined by finding the 'starter cells' from which the retrograde transsynaptic spread of the barcoded rabies virus originates. As this technique determines connectivity at the

resolution of individual cells, transcriptomic information can also be determined for each of the barcoded cells, enabling the study of connectivity motifs between transcriptomic cell types.

I designed my barcoded rabies constructs to be used with this monosynaptic tracing strategy, to label each starter neuron with a random nucleotide barcode sequence that is expressed by each of the rabies viral particles from a library of diversely barcoded viruses. As discussed in the introduction of Chapter 2, the barcode diversity of this library affects the ability to determine the connectivity of neurons with these barcodes. If the library is dominated by a few abundant barcode sequences that are found in most of the viral particles, then the probability of labelling multiple starter neurons with the same barcode increases. If two starter cells share the same barcode, so will each of their presynaptic cells that the rabies spreads to. These presynaptic partners cannot be unambiguously mapped to either of the starter cells, so this connectivity information is lost. The diversity of my barcoded libraries is sufficient for this to be a rare event as long as the total number of starter cells in an experiment is kept below a few thousand cells (estimated 93.0% uniquely labelled cells for 2000 starter infections, Figure 2.6).

However, this is not the only way that starter cells may share a barcode. In highly connected neural circuits it is possible that rabies will spread from one starter neuron to another starter neuron. As there is rabies glycoprotein present in these starter neurons, the rabies virus could spread to presynaptic neurons from the initially infected starter cell or the secondarily infected starter. Regardless of whether spread occurs from this second starter, when we observe the barcodes from the other presynaptic cells we have no way to tell which of the two starters these barcodes originated from. In an idealised experiment, we could choose to exclude any barcodes that we find in multiple starter cells from further analysis so that this ambiguity is removed from the connectivity we observe. However, in an actual experiment there are multiple reasons why we might not detect all of the cells which have been infected with rabies barcodes. These possibilities can include loss of tissue due to physical damage, inefficiencies in the detection of rabies barcode sequences or the detection of markers of starter cells, and the death of neurons due to rabies virus infection. These missed detections could result in losing a barcoded starter cell which shares a barcode with another starter cell, resulting in the remaining starter cell being mistakenly identified as being uniquely barcoded. Therefore, it is necessary to minimise the likelihood of transsynaptic spread between starters whilst simultaneously minimising the number of undetected barcoded cells in an experiment, so that the observed barcodes can be relied upon as an accurate representation of neuronal connectivity.

In this chapter, I will outline my experimental approach for reducing starter-to-starter barcode spread, with the aim of unambiguously determining the connectivity of neurons with barcoded rabies viruses. I will then demonstrate the application of this barcoding technique in the mouse visual cortex and present analyses of the putative connectivity

data from this dataset. I will also describe what the resultant cell-barcode information tells us about the conditions that I achieved in this experiment and what can be further improved to use barcoded rabies viruses as a reliable tool for determining connectivity.

5.3 Methods

5.3.1 Animals and surgical procedures:

All animal research was conducted in compliance with the guidance and regulations set forth by the UK Home Office according to the Animals Scientific Procedures Act 1986. The surgeries performed were approved by the Named Veterinary Surgeon at the Francis Crick Institute under project licence PP4882546. Surgical procedures were carried out on male and female C57BL/6 mice aged between 12 and 15 weeks.

Surgical procedure for mice used in barcoded rabies sequencing experiments

Prior to anaesthesia, mice received a 50 µl tail vein injection of AAV-PHP.eB-Syn-Cre (4.5 x 10^{13} vg/ml) diluted to 1:330 in sterile saline. Mice received Metacam peri-operative analgesia via water before surgery (5 mg/kg). Mice were anaesthetised using 1.5-3% isoflurane and received a subcutaneous injection of Buprenorphine (0.1 mg/kg) and Metacam (10 mg/kg). Mice were secured in a stereotactic frame, and a small craniotomy was created at the site of the injection. A Nanoject III injector was utilized to deliver viruses at a speed of less than 14nl per minute. The initial injections of helper viruses AAV1-FLEX-nuc-mCherry-N2cG (3.1 x 10^{13} vg/ml) and AAV-FLEX-nucEBFP2-TVA (1.8 x 10^{13} vg/ml), diluted to 1:20 and 1:100 respectively, were delivered at three injection sites with coordinates of 3.5mm AP and 3.0, 2.4 and 1.8 mm ML, at a depth of 300 and 500 µm per injection site (60nl per depth). Seven days later, a pair of 500nl injections of barcoded RV35 CVS-N2c-mCherry (6.0 x 10^{8} IFU/ml) were administered at the same coordinates. Mice received further post-operative analgesia of Metacam orally for three days after surgery (5 mg/kg). Seven days later, to extract the brains for in situ sequencing mice were euthanised with sodium pentobarbital. The brains were extracted without prior perfusion and flash-frozen in OCT-filled moulds using a bath of absolute ethanol and dry ice, and subsequently stored at -80°C.

Surgical procedure for mice used in control injections, PHP.eB viral titration, and serial two-photon tomography experiments

For TVA leak and rabies glycoprotein-coated contamination control experiments, procedures were performed exactly as above but without prior tail vein injections, and without AAV1-FLEX-nuc-mCherry-N2cG injection and/or without AAV-FLEX-nucEBFP2-TVA

injection. When extracting the brains, the mice were euthanised with sodium pentobarbital and perfused with ice-cold PBS followed by 4% PFA in PBS. Brains were then further fixed in 4% PFA overnight. For AAV-PHP.eB-Cre titration experiments, mice received a 50 µl tail vein injection of AAV-PHP.eB-Syn-Cre (4.5 x 10^{13} vg/ml) diluted to 1:100/1:330/1:1000/1:3300 in sterile saline. Two weeks after tail vein injections, mice were euthanised with sodium pentobarbital. The brains were perfused and fixed as above. For barcoded rabies serial two-photon tomography, the procedure for barcoded rabies sequencing described above was followed with the following differences: a single local injection of AAV1-hSyn-Cre (7.6 x 10^{13} vg/ml) diluted 1:5000 in sterile saline was mixed with the helper viruses diluted as above and injected at coordinates of 3.5mm AP, 2.5mm ML, at a depth of 300/500 µm (60nl per depth). Seven days later, a pair of 500nl injections of barcoded R2 CVS-N2c-mCherry (2.8 x 10^{7} IFU/ml) were administered at the same stereotactic coordinates. After another seven days, mice were euthanised with sodium pentobarbital and perfused and fixed as described above.

5.3.2 Serial section two-photon tomography

After fixation, brains were transferred to PBS and if necessary, were kept long-term in PBS with 0.01% Sodium Azide. Brains were then transferred to a serial sectioning two-photon microscope (Mayerich et al., 2008; Ragan et al., 2012). Sections were imaged in a bath of 50 mM pH7.4 Phosphate buffer. The microscope was controlled by ScanImage Basic (MBF Bioscience), using the custom software wrapper BakingTray to control imaging parameters (Campbell, 2020). Image tiles were assembled by StitchIt (Campbell et al., 2020). Images for cell counting of AAV PHP.eB Cre positive cells were acquired at 2 µm XY resolution with 40 µm vibratome sections, imaged with optical sectioning at 8 µm Z resolution. Images for cell counting of rabies mCherry positive cells were acquired at a higher resolution of 1x1 µm XY pixels with 5 µm z-spacing.

The cellfinder software pipeline (Tyson et al., 2021) was used to detect cells across the brain. The pre-trained deep neural network was finetuned with extra example cells from our microscope setup to improve false positive and false negative performance using the napari interactive plugin by manually annotating true cells and cell-like background signal.

5.3.3 In situ sequencing experiments:

Tissue treatment

The tissue treatment steps described in Chapter 4 (Methods 4.3.3) were completed with the following additions and alterations:

Reverse transcription

During the reverse transcription step (Methods 4.3.3), an additional rabies-specific reverse transcription primer containing a series of 5' locked nucleic acid bases for increased template binding specificity and sensitivity was added to the RT mix (RTAB0840, Appendix Table 7.4).

Padlock ligation

After the gene padlock ligation step had been carried out, a second gap-filling/ligation mix was made, containing: 1 µM rabies barcode padlock (OAB0973, Appendix Table 7.4), 0.4U/µl RNaseH (New England Biolabs, M0297L), 20% Formamide, 50 mM KCl, 0.25 µg/µl non-acetylated BSA (Thermo Fisher Scientific, AM2616), 50 µM dNTP (Thermo Fisher Scientific, R0192), 5% glycerol, 0.001U/µl Phusion polymerase (Thermo Fisher Scientific, F530L), 0.5U/µl Ampligase (Lucigen, A0102K) in 1X Ampligase buffer. The samples were incubated in the gap-filling reaction mix for 5 minutes at 37°C, followed by 40 minutes at 45°C. Then the samples were washed twice in PBST and once in hybridisation mix and the rolling circle amplification step proceeded.

Rolling circle amplification

During the addition of the RCA primer mix, an extra RCA primer specific to the rabies padlock (PAB0035, Appendix Table 7.4) was added to the mix at a final concentration of 1 µM. Hybridisation and RCA then proceeded as described in Chapter 4 Methods 4.3.3.

mCherry imaging and photobleaching

After rolony production was complete, rather than proceeding immediately to gene sequencing, the samples were first imaged to detect the presence of mCherry-positive helper virus-infected cells using 500 nm excitation light and the same dichroic and emission filters used for sequencing rounds. After imaging, the samples were placed on an aluminium block under a white-light 10,000-lumen LED floodlamp (NATPOW, TS-F42100) overnight to photobleach the mCherry signal so that it didn't interfere with subsequent sequencing. A fan was used to maintain a consistent sample temperature of under 35°C during photobleaching. The following day, gene sequencing proceeded as described in Chapter 4 Methods 4.3.3.

Rabies barcode sequencing

Sequencing of the rabies barcode took place directly after the 7 rounds of gene sequencing were complete, before the hybridisation rounds. A stripping buffer of 60% Formamide in

2X SSC buffer was added to each flowcell and incubated at 60°C for 15 minutes. This was then repeated, adding fresh stripping buffer for another 15 minutes. The samples were then equilibrated in hybridisation mix. Then 1 µM rabies barcode sequencing primer (PAB0034, Appendix Table 7.4) in hybridisation mix was added to the sample and incubated at room temperature for 1 hour. Excess sequencing primer was washed away with three washes of hybridisation mix at room temperature for two minutes each. Then samples were washed twice in PBST before proceeding to sequencing. The same sequencing procedure as that for gene barcodes (Chapter 4 Methods 4.3.3) was carried out for rabies barcodes, completing up to 14 rounds of sequencing for the rabies barcode. After rabies barcode sequencing, the rabies sequencing primer and its extended barcode sequence were stripped from the samples in the same way as described for the gene sequencing primer, and hybridisation rounds proceeded.

5.3.4 Image processing and rabies barcode detection

The image processing steps of: image projection; registration between channels and rounds; image registration between modalities; tile stitching within ROIs and registration to the Allen Atlas Common Coordinate Framework, were performed in the same manner as described in Chapter 4, Methods 4.3.4 for gene sequencing round and rabies barcode sequencing rounds. Rabies barcode rolonies were then extracted from each imaging tile with the following method:

Base calling rabies barcodes

A set of manually selected tiles that contain rabies rolony signals are convolved with a difference of Hanning kernel with a radius that is the size of an average rolony, removing high-frequency noise. Normalisation factors are then calculated per-channel and per-round by computing the distribution of pixel intensities across all channels and rounds. Linear regression is used to fit the signal drop-off across rounds. Next, the locations of isolated rolony spots are detected from the set of manually selected tiles that contain rabies rolony signals. This is done by finding peaks (spots) in the image and then computing the average intensity in an annulus around each detected spot, selecting spots that fall below an isolation threshold. The images are normalised across rounds to the intensity of the first round using the normalisation factors. Then the fluorescence values across channels and rounds are extracted from disks with a 2 pixel / 0.462 µm radius in the locations of these spots.

Scaled K-means clustering (Bugeon et al., 2022) is used to define each spot as belonging to one of four clusters corresponding to the signal from the four dye-conjugated nucleotides. The cluster means of the four clusters are computed across barcode rounds. Then for each tile, the barcode spots are detected from the registered round/channel images by

finding peaks in the mean image across round and channels and then comparing each found peak's shape against an average spot image shape with a positive inner radius and a negative outer ring to penalise elongated non-spot like shapes. From each spot location that passes the shape similarity threshold to the average spot image, the fluorescence trace across rounds and channels is extracted from a disk with a 2 pixel / 0.462 µm radius.

Cluster mean fluorescence intensities and spot traces are normalized to unit norm across channels. Normalizing across channels ensures that comparisons between the traces focus on their relative patterns (e.g. distribution across channels) rather than the absolute intensities of the channels, which can be affected by differing excitation intensity across dyes and the quantum efficiency of the dyes themselves. For each imaging round, the cosine similarity scores are computed between channel normalized spot traces and the channel normalised cluster means of each round, assigning each spot to the cluster with the highest score. Unassigned bases from missing data are labelled as "N." Barcode sequences are constructed from these assignments. Quality metrics are then calculated, including the mean fluorescence intensity, the mean cosine similarity score across rounds, and a separate cosine similarity score of the whole sequence, calculated by comparing the observed fluorescence traces (after background subtraction) to synthetic traces generated from the cluster means of the bases which were assigned each round.

Rabies barcode quality control and error correction

After base-calling rabies barcode sequences within each tile, all of the detected barcode spots are aggregated across the whole experiment, retaining their ROI and coordinate positions. First, thresholds are applied on the mean intensity score, cosine similarity score and mean score to remove bright autofluorescent spots which do not change colour between rounds. To further filter the remaining spots, a Gaussian Mixture Model (GMM) is applied to the four metrics: cosine similarity score, spot score, mean intensity, and mean score. A subset of the data is sampled by selecting every 10,000th spot and initial cluster means are set using percentile values of the sampled data, ranging from low (1st percentile) to high (99th percentile). The GMM is configured with two clusters, representing low and high metric values, and is fitted to the sampled data. The model then classifies all barcode spots into two clusters, assigning a label to each spot based on its metrics. Spots assigned to the high-quality cluster are retained for further analysis, while low-quality spots are excluded.

Error correction is performed on the barcode sequences from the remaining set of high-quality barcode spots using an algorithm that identifies and merges sequences within a specified maximum hamming distance. Unique sequences are then identified, counted, and sorted by abundance. For each unique sequence, the hamming distance to all the lower-abundance sequences is calculated, and those within the specified threshold (hamming distance = 2, minimum matching bases = 7) are corrected to the high-abundance sequence.

Rabies barcode assignment to cell masks

Each spot represents a barcode that needs to be assigned to one of the available masks or to the background. We have some prior expectations of the data which we have used to create a model for assigning these spots to cells. First, we know that infections of a cell will result in the replication of a rabies barcode and that we should expect multiple spots in close proximity to each other and to the centre of a cell if the cell is infected. Secondly, we know that barcodes can be detected in the neuronal processes which are present throughout the tissue, so sparse barcode spots are expected to be detected everywhere. We cannot confidently map these to a local cell in isolation so we must have a background assignment for these.

The problem of assigning spots to cells is cast as a form of iterative likelihood maximisation by shifting spot assignments after an initial assignment step, using a Gaussian spatial likelihood function which considers the distance of a spot to a mask, a background component with a fixed low-density likelihood to account for spots that do not correspond to any specific mask, and a mask spot-count likelihood which penalises spot assignments to cells unless justified by the spatial data.

We assume that each spot $\mathbf{x}_i$ can be assigned to either a mask $j \in \{1, \ldots, M\}$ or to the background (denoted by -1).

Mask assignment likelihood (Gaussian)

$$\ell(\mathbf{x}_i, j) \;=\; \exp\Big(-\frac{1}{2\,\sigma^2}\|\mathbf{x}_i - \boldsymbol{\mu}_j\|^2\Big), \quad \text{for } j = 1, \ldots, M.$$

where $\mathbf{x}_i \in \mathbb{R}^2$ is the 2D coordinates of spot i, $\boldsymbol{\mu}_j \in \mathbb{R}^2$ is the centroid of mask j, and σ is the Gaussian distribution scale parameter.

Background assignment likelihood

$$\ell(\mathbf{x}_i, -1) \;=\; \alpha$$

where α is a small constant (e.g. 10^{-6}) representing the background likelihood.

Mask spot-count likelihood We introduce a mask spot-count likelihood factor (5.1):

$$\ell(n_j(\mathbf{z})) \;\propto\; \exp\Big[-\frac{(n_j(\mathbf{z}))^{\beta}}{m}\Big],$$

where $n_j(\mathbf{z})$ is the number of spots assigned to mask j under the assignment vector $\mathbf{z}$, m is a scale factor, and β is an exponent that controls how the likelihood penalty scales with spot count. Setting $\beta < 1$ makes the penalty sublinear in n_j, so the incremental penalty for each additional spot decreases once a mask already has some spots.

Posterior probability Combining the spot-wise assignment likelihoods with the spot-

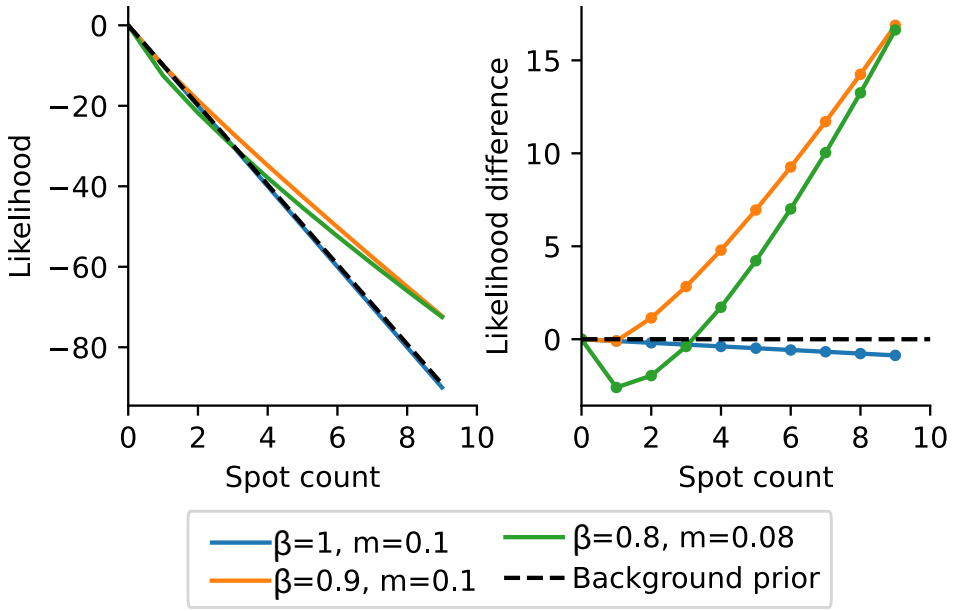


Figure 5.1: Contribution of mask spot count likelihood with increasing spots. Plots demonstrating the change in resultant likelihood with increasing barcode spot count with different parameters of the mask spot-count likelihood, calculated without contributions of the Gaussian distance likelihood. Background prior in dashed black line as the reference for background/mask assignment.

count likelihood term, the posterior for the whole assignment vector $\mathbf{z}$ is:

$$P(\mathbf{z} \mid \{\mathbf{x}_i\}, \{\boldsymbol{\mu}_j\}) \;\propto\; \prod_{i=1}^{N} \ell(\mathbf{x}_i,\, z_i) \;\times\; \prod_{j=1}^{M} \ell(n_j(\mathbf{z})).$$

Where N is the total number of spots and M is the total number of cell masks.

Each unique barcode sequence is handled in isolation, meaning that the assignment logic, likelihood calculations, and prior distributions applied to one barcode do not affect the assignments of other barcodes. Initially, spots are assigned to the nearest mask based on Euclidean distance, provided they fall within a specified maximum distance threshold. Spots that do not meet this criterion are assigned to the background.

From this point of initialisation, the algorithm works by an iterative optimisation process designed to refine the initial assignments. In each iteration, the algorithm explores potential reassignments of spots to masks by evaluating combinations of spots that are found close together, defined by an inter-spot distance threshold. The algorithm generates all valid combinations of nearby spots up to a maximum group size, ensuring computational feasibility by limiting the total number of combinations considered. For each valid combination of spots, the algorithm calculates the change in likelihood that would result from reassigning the group to a different mask or to the background. The algorithm prioritizes reassignments that yield the highest increase in overall likelihood, thereby iteratively improving the accuracy of spot-to-mask assignments. On each iteration, the algorithm evaluates two scenarios for each spot combination: moving the spots to a new mask and

assigning them to the background.

Upon identifying the optimal reassignment that maximizes the likelihood, the algorithm updates the mask assignments accordingly. This process is repeated iteratively, with each iteration potentially involving multiple spot reassignments across different masks. The algorithm tracks the movement of spots to prevent conflicting assignments within the same iteration (e.g. the same spot in two different groups being assigned to two different masks) and assignments with the highest likelihood are prioritised. The iterative optimisation continues until one of the following conditions is met: no further spot movements result in an increased likelihood, or the algorithm reaches a predefined maximum number of iterations. In cases where no spots are moved in an iteration, the algorithm attempts to assign all remaining unassigned spots to the background, further refining the assignment. The final output is a comprehensive assignment of each spot to a specific mask or to the background.

mCherry cell detection

To robustly detect the nuclear mCherry fluorescence signal of helper-virus-infected cells and prevent erroneous detection of autofluorescence, a linear regression model is fitted to a channel containing only background autofluorescence and the channel containing mixed background and signal data. The regression coefficients obtained are used to predict and subtract the background contribution from the mixed signal image, yielding the isolated pure mCherry signal image. A difference of Gaussians filter is applied to the unmixed mCherry image to remove out-of-focus halos from the epifluorescence imaging of these bright objects. A binarised version of the image is created with a set binarisation threshold and binary closing is performed. The properties of each putative cell mask such as area, eccentricity, circularity, solidity, and intensity of pixel values from the unbinarised region, are calculated. Then the masks are filtered based on minimum and maximum area, circularity, elongation, solidity and intensity thresholds. For mCherry masks in the overlap region between two tiles, overlapping masks are searched for. If the overlap proportion for a pair of masks between the two tiles exceeds a predefined upper threshold (30%), the label in the shifted tile is removed to resolve significant overlaps. Conversely, if the overlap proportion falls below the threshold, the shared region is excised from both masks to address minor overlaps. This prevents multiple instances of the same infected cell from being detected.

mCherry assignment to cell masks

To assign mCherry masks to rabies-infected cells, all of the reference-aligned tile coordinates are used to make full ROI stitched images of the mCherry masks and rabies spots coordinates. Each mCherry mask is first checked for any rabies spots which overlap with

it. If barcode spots are found in the mCherry mask, the cell ID of the most abundant barcode is checked to find all of the total barcodes present in that rabies-infected cell. If the percentage of spots within the mCherry mask is above a threshold (conservatively 10% of the total spots in the rabies-infected cell), and the centroid of the spots overlaps with the mCherry mask, then the mCherry mask is assigned to the cell mask corresponding to the barcode spots.

5.4 Results

5.4.1 Titrating starter cell density *in vivo*

Before directly applying barcoded rabies viruses to determine connectivity, I sought to define a set of experimental conditions for viral injections that would ensure that the prevalence of transsynaptic spread of barcodes between starter cells was minimal. If the connection probability between a pair of cells was invariant to the distance between the cells, then the only way to achieve this would be to reduce the absolute number of starter cells that are present in a single experiment. However, it has been demonstrated in the mouse and rat cortex that connection probability decreases with increasing intersomatic distance between pairs of neurons (Holmgren et al., 2003; Perin et al., 2011; Song et al., 2005), and this relationship can be modelled by a Gaussian fit with parameters dependent on cell types (Campagnola et al., 2022; Levy and Reyes, 2012). Therefore, increasing the average distance between starter cells should reduce the probability of starter-starter connectivity. To maintain a high number of starter cells in a single animal for increased experimental throughput, the starter cell density must be decreased, spreading the infected cells over a larger area.

A starter cell is defined by the expression of two proteins: TVA, which permits initial pseudotyped EnvA-RVΔG infection, and rabies glycoprotein, which enables the rabies virus to spread transsynaptically. In my experiments, I induced the expression of these proteins by using two helper AAV viruses which are both Cre-dependent: AAV1-FLEX-nuc-mCherry-N2cG and AAV-FLEX-nucEBFP2-TVA. In order to achieve a uniform, sparse density of Cre-positive neurons to seed starter cells, I used the PHP.eB engineered capsid variant of AAV9 which is able to efficiently cross the blood-brain barrier (Chatterjee et al., 2022) and deliver a Syn-Cre construct uniformly across the CNS via tail vein viral delivery.

A series of viral dilutions of AAV-PHP.eB-Syn-Cre were injected into Ai14-tdTomato Cre-reporter mice. After 3 weeks, the brains were extracted and imaged using a serial-sectioning two-photon microscope to detect tdTomato-positive cells. The brains were then registered to the Allen reference atlas and cells were automatically detected throughout each brain volume using cellfinder (Methods 5.3.2). Cell densities were calculated for each brain area and mean densities were calculated for each dilution cohort (Figure 5.2).

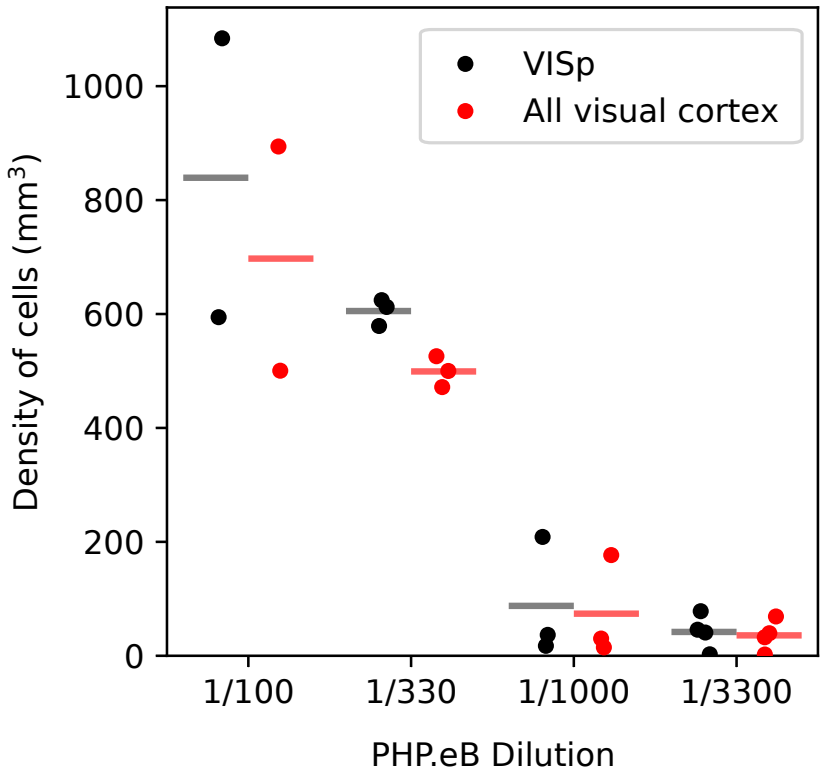


Figure 5.2: Dilution of AAV-PHP.eB can titrate Cre-positive cell density. Density of detected tdTomato cells in Ai14(RCL-tdT)-D Cre reporter mice after tail vein injections of diluted AAV-PHP.eB-Cre in VISp (primary visual cortex) - black, and all of visual cortex - red. Bars represent mean density (N = 4 mice for 1/3300, N = 3 mice for 1/1000, 1/330, N=2 for 1/100 dilution).

As expected from previous characterisations of the PHP.eB serotype (Chatterjee et al., 2022) the virus was able to induce CNS infection even at low titres, and dilution of the virus correlates with the observed density of infected cells, enabling dilution-based control of starter density. However, variability of observed infection density is present in both high (1/100) and low titre (1/1000) injections, possibly due to the tail-vein route of administration.

A second factor which influences the probability of transsynaptic spread between starter cells is the degree of transsynaptic labelling which occurs per starter cell. To determine the number of presynaptic cells per starter, mice were injected with AAV1-hSyn-Cre, AAV1-FLEX-nuc-mCherry-N2cG and AAV-FLEX-nucEBFP2-TVA in the visual cortex and then 14 days later were injected with EnvA-CVS-N2c-mCherry-BC barcoded rabies. 7 days after that, brains were extracted and imaged in the same way as the PHP.eB cohort (Figure 5.3 A, B). Brains were registered to the Allen atlas and all mCherry-positive rabies cells were counted, as with the PHP.eB brains. As expected, cell counts were highest in the injection site of the primary visual cortex, followed by the higher visual areas, lateral posterior thalamus and the lateral geniculate nucleus (Figure 5.3 C). By cell density, areas of the lateral geniculate nucleus had the highest values, followed by the primary visual cortex and higher visual areas (Figure 5.3 D). The total number of presynaptic cells found

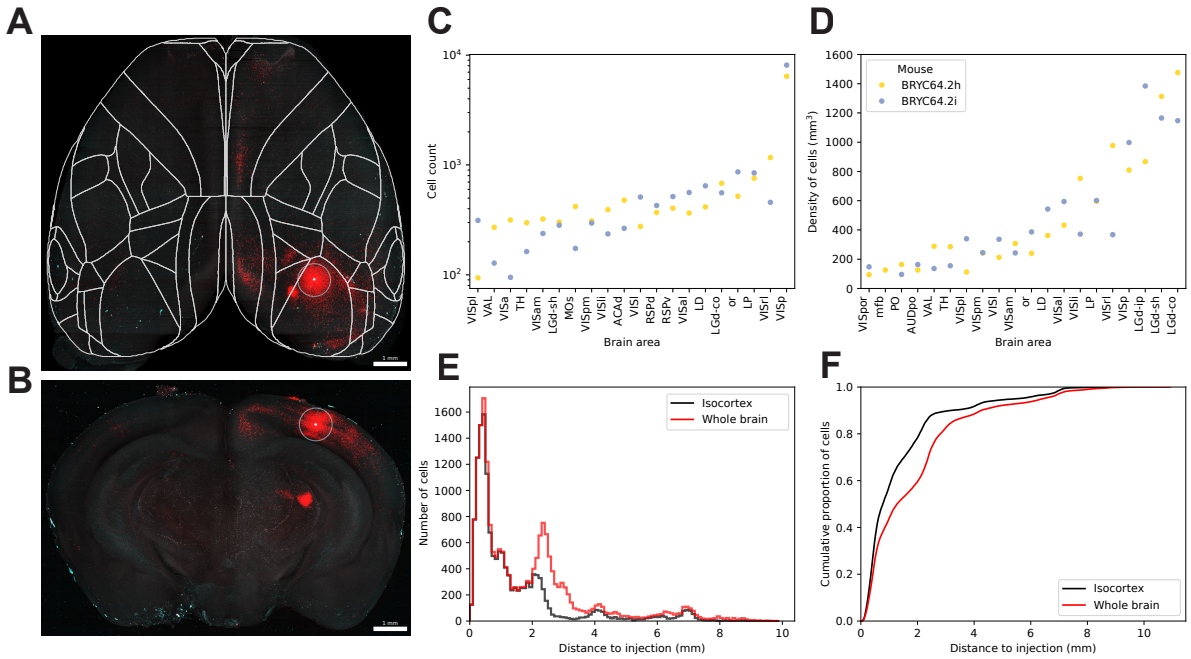


Figure 5.3: Quantifying spatial distribution of transsynaptic rabies spread in the whole brain. **(A)** Dorsal view of the cortex of a brain injected with barcoded rabies virus expressing mCherry (red) (grey circle - injection site centre) with an overlay of cortical area boundaries (white), with a coronal view, **(B)** of the same brain at the centre of the injection site (grey circle - injection site centre). **(C)** Counts of detected rabies-positive cells in the 20 highest abundance regions of the brain. **(D)** Density of detected cells per mm3 in the 20 highest density brain regions. **(E)** Distribution of rabies cell distances to the centre of the injection site (cells detected throughout the whole brain - red, cells detected within just the isocortex - black) and **(F)**, the corresponding cumulative distribution of those distances.

in the two brains were 17,505 and 19,139, with a visible peak in the distribution of the cell's distance from the centre of the injection site at 2 mm, demarking the infection in the lateral geniculate nucleus (Figure 5.3 E, F).

It was possible to distinguish the nuclear-mCherry signal of glycoprotein-positive starter cells from the cytoplasmic rabies fluorescence in optical sections of the images from the serial-sectioning two-photon microscope (Figure 5.4 A, B), but to ensure that all of the starters from the injection site were counted, all of the brain sections were retrieved and mounted onto slides for confocal imaging (Figure 5.4 C, D). Manual cell counting of starter cells gave a total of 263 cells. This gives an average of 67 presynaptic cells per starter cell across the entire brain. An estimate of the probability of transsynaptic spread between starter cells was then made using the observed values of presynaptic cell counts per starter and the density of Cre-positive neurons. The following calculation provides a rough approximation of the expected probability of spread, in order to select a dilution of AAV-PHP.eB-Cre to proceed with for barcoded rabies connectomic experiments. Assuming that presynaptic cells are infected at random from all cells within the injection site, the probability of transsynaptic spread between starter cells is given by:

$$P_{\text{transsynaptic starters}} = 1 - (1 - P_{\text{starter}})^{N}$$

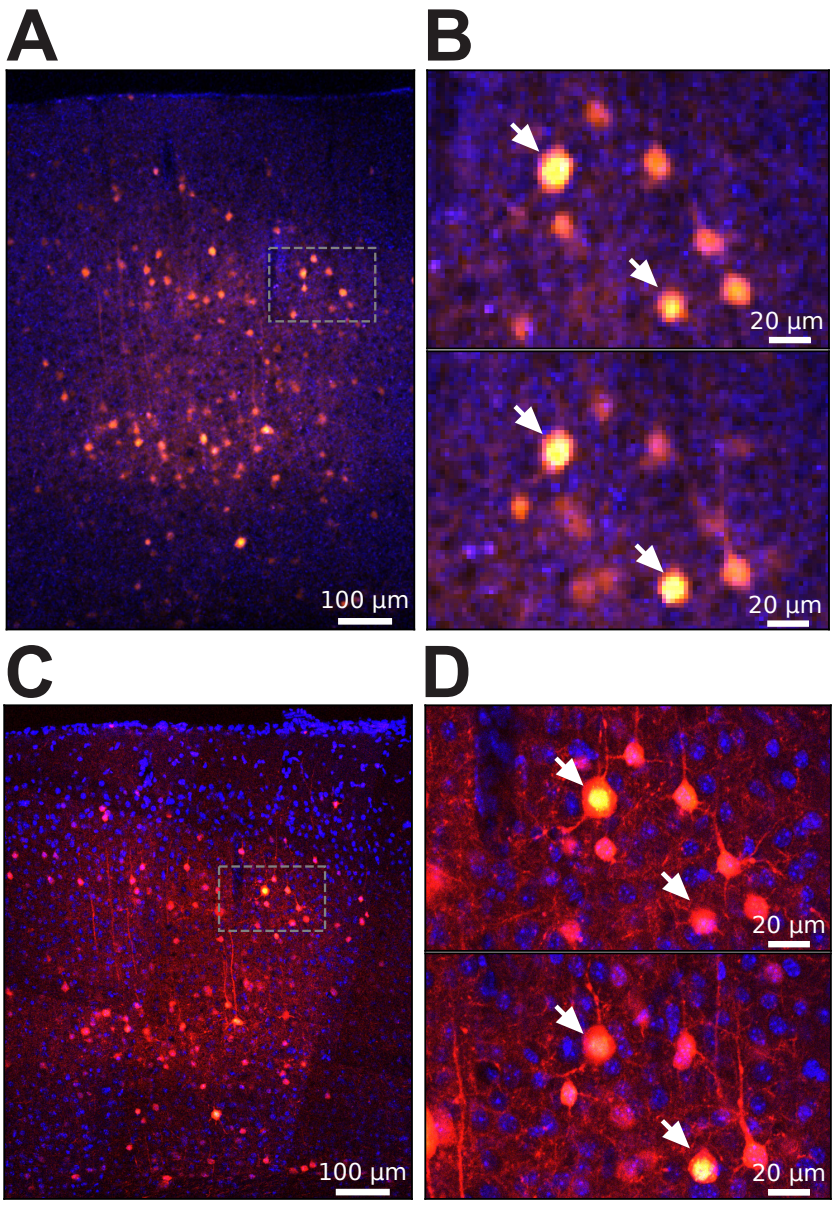


Figure 5.4: Glycoprotein-positive starter cells can be counted after serial two-photon imaging. **(A)** Serial two-photon images of a coronal section of the injection site of a barcoded rabies infected brain (rabies mCherry - red, autofluorescence channel - blue). **(B)** Two optical imaging planes with detected helper virus infected cells expressing nuclear-mCherry (white arrows). **(C)** A matched confocal image of the same coronal section, post-stained with DAPI (blue). **(D)** Two z-planes showing the nuclear mCherry expressing cells (white arrows).

where P_{starter} is the probability that a cell from the injection site is a starter cell and N is the number of presynaptic neurons per starter that are found in the injection site. I determined the number of presynaptic cells that are found within 1 mm of the centre of the injection site to be 20.9 presynaptic cells per starter (5512 presynaptic cells / 263 starter cells). A distance of 1 mm was chosen based on the maximum observed infection distance from the centre of a helper AAV injection site using the same titres of helper viruses that I use for transsynaptic rabies experiments (see Methods 5.3.1 for titres and volumes used in transsynaptic experiments). The probability of a cell being defined as a starter was calculated for each PHP.eB dilution by dividing the mean density

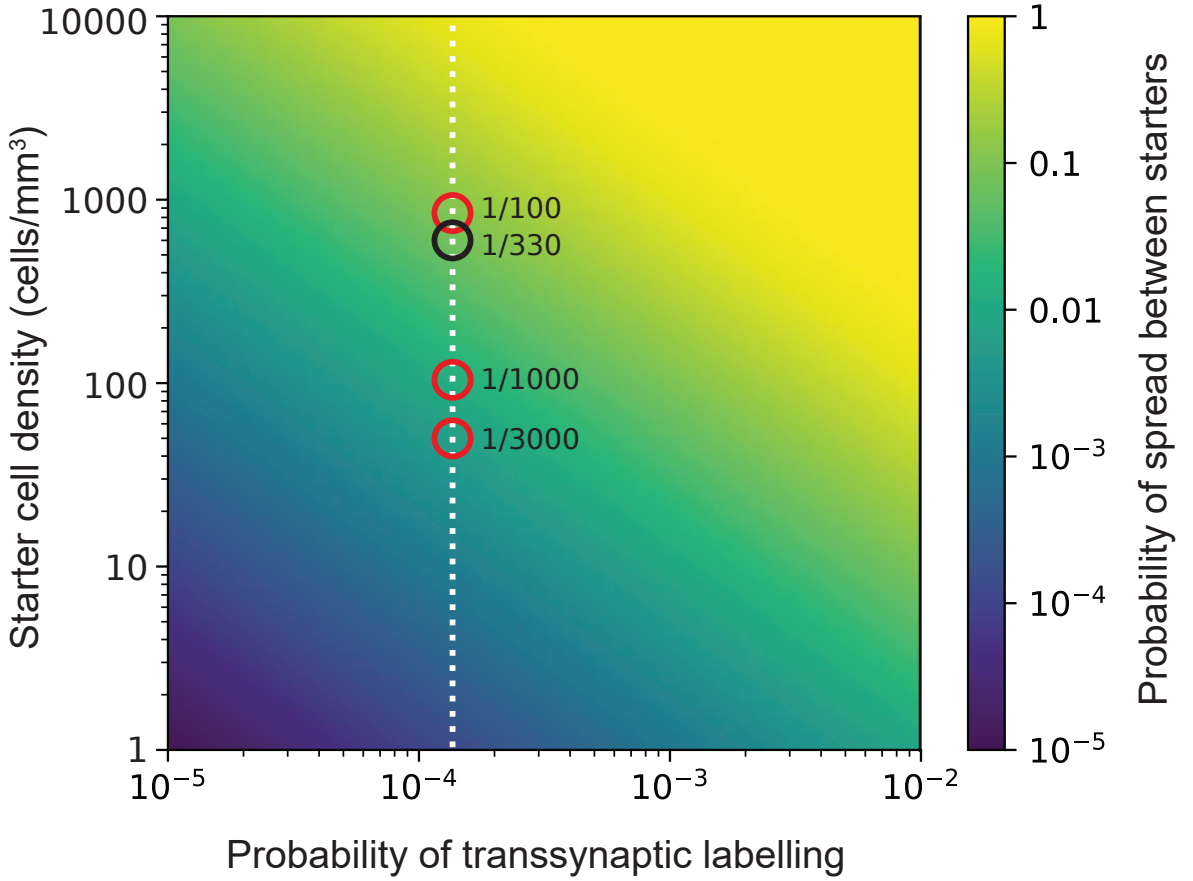


Figure 5.5: Estimation of transsynaptic spread between starter cells. A heatmap of the predicted probability of transsynaptic spread between starter cells given the probability of a cell being transsynaptically labelled by a starter cell (x-axis) and the density of starter cells (y-axis). The estimate of transsynaptic labelling probability (1.39×10^{-4}) is denoted with a dotted white line, with the mean cell density of each AAV-PHP.eB-Cre dilution shown as a circle along the line.

of Cre-positive cells detected in the visual cortex by the average density of neurons in the visual cortex (150,000 cells/mm^3, Herculano-Houzel et al., 2013; Murakami et al., 2018). The resultant probability of observing transsynaptic spread from starter to starter is plotted for each dilution (Figure 5.5). The 1/330 dilution was selected for further barcoded rabies experiments with the assumption that the 8.1% estimated probability of transsynaptic spread between starters for this dilution represents an upper limit if all of the Cre-positive neurons in the injection site are infected by the glycoprotein AAV. This doesn't take into account any biases in rates of connectivity between different cell types, or the fact that connection probability reduces with distance and was used as a starting point to be refined empirically.

Different AAV serotypes have been found to have distinct tropisms within the CNS (Shen et al., 2011; Summerford and Samulski, 1998; Wu et al., 2006). Recent scRNAseq (Brown et al., 2021) and spatial transcriptomic analyses (Jang et al., 2023) of AAV-PHP.eB tropism has shown it to exhibit a bias towards inhibitory neurons and layer 5 excitatory neurons. Using the detected cell locations from the whole brain imaging of PHP.eB dilu-

tions, I quantified the distribution of infected cells across layers of the cortex. Throughout the volume of the primary visual cortex and of the whole isocortex, the largest percentage of cells were found in layer 5, followed by layer 2/3 (Figure 5.6 A, C). This does not account for the thickness or density of total cells within each layer but shows that a large proportion of starter cells are likely to originate in these layers when using AAV-PHP.eB. When accounting for the volume of each layer, the density per layer is still highest in layer 5 for the primary visual cortex, followed by layer 4 and layer 6b. Density across the whole cortex is highest in layer 4, then layer 5 (Figure 5.6 B).

5.4.2 In situ sequencing of barcoded cells

I proceeded to perform readout of rabies barcodes and endogenous genes using the in situ sequencing approach detailed in Chapters 3 and 4. A 1/330 dilution of AAV-PHP.eB Cre was combined with three local injections of the helper AAVs AAV1-FLEX-nuc-mCherry-N2cG and AAV-FLEX-nucEBFP2-TVA performed in a coronal plane within the primary visual cortex (Methods 5.3.1). The AAV-PHP.eB and local AAV injections were performed on the same day, as it was found that performing local AAV injections 2 weeks after AAV-PHP.eB tail vein injection resulted in inflammation of the brain surface of the injection site, likely as the result of an immune response at the surface of the brain, provoked by the second AAV injections after the initial systemic insult from the AAV-PHP.eB. 2 weeks after helper virus injection, the barcoded rabies virus library - RV35 was injected at the same sites as the AAV helper viruses. The brain was then extracted and flash-frozen in preparation for cryosectioning and sequencing. All of the following data in this chapter, unless otherwise noted is derived from a series of 40 serial coronal cryosections of 20 µm thickness covering 800 µm of tissue centred on the injection sites of the barcoded rabies virus (Methods 4.3.3). The extent of the brain coverage of these cryosections can be observed from the flatmap projections of Figure 4.13 and the coronal projections of Figure 5.16 A and B. Sequenced tissue covered the anterior primary visual cortex, the higher visual areas: posteromedial, anterolateral, lateral, laterointermediate visual cortex, as well as retrosplenial cortex, auditory cortex, ectorhinal area, temporal association areas and subcortical structures including the thalamus. Importantly, while the sequenced cryosections cover almost all of the injection site, one limitation of this dataset is that a fifth flowcell that was furthest anterior from the injection site still contained some mCherry-positive neurons, which may have been starter cells. Due to technical reasons, gene and barcode transcripts from this flowcell were not efficiently detected, so the full contribution of the putative starter cells from this flowcell cannot be assessed. Across the 40 cryosections which were sequenced in the other flowcells, 4122 mCherry nuclei were detected. Of those, 575 were detected as rabies positive. Within the missing flowcell, I conservatively estimate the total number of mCherry nuclei to be 100 from manual counting. Given the average rate of mCherry infection, this would mean the loss of 14

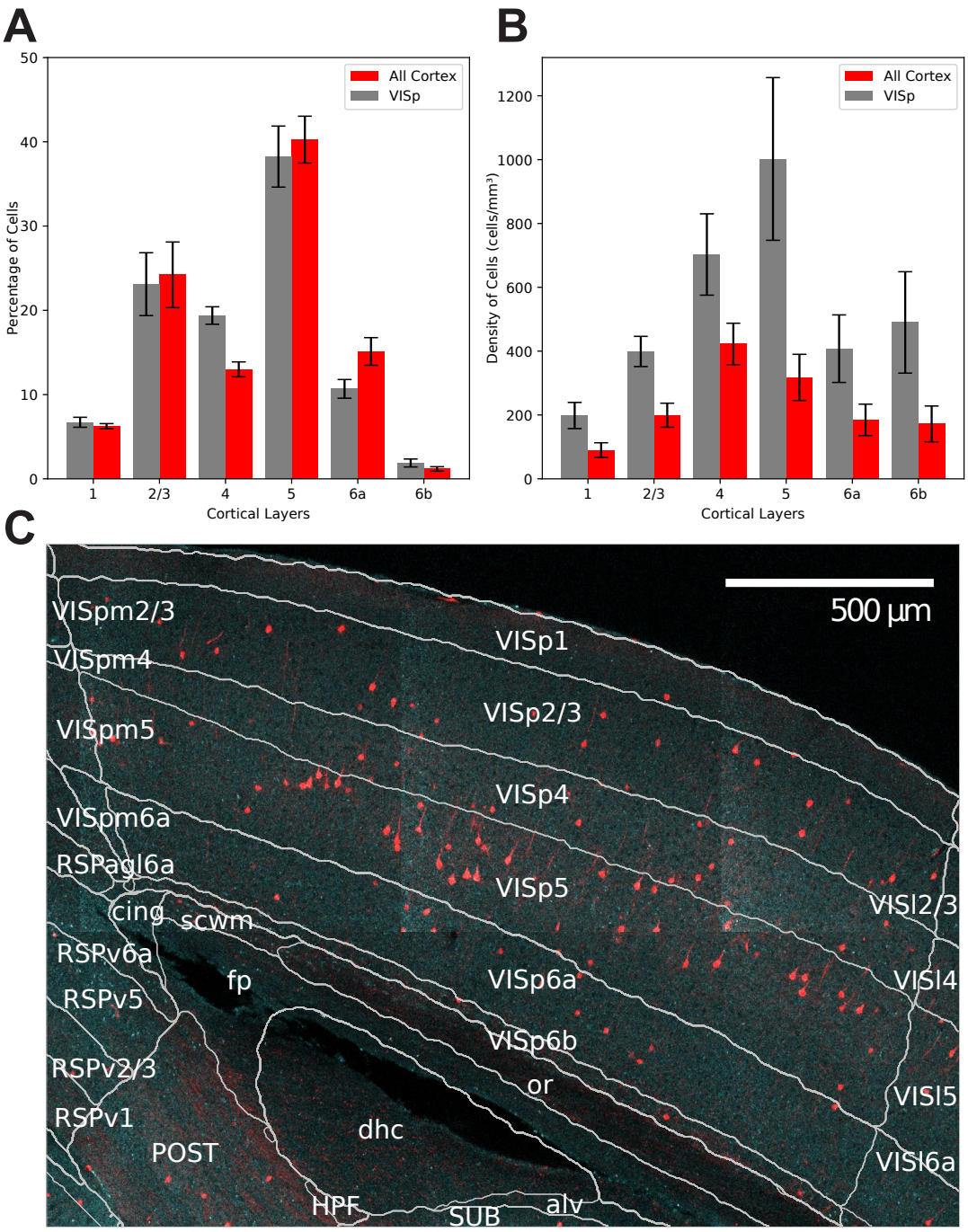


Figure 5.6: AAV-PHP.eB shows layer-specific biases of infection in visual cortex. **(A)** Distribution of detected tdTomato cells in Ai14(RCL-tdT)-D Cre reporter mice after tail vein injections of 1/330 dilution AAV-PHP.eB-Cre. Bars represent percentage of total cells from primary visual cortex (grey) and all of the isocortex (red) found in cortical layers. **(B)** Same as **(A)** with density of detected cells per layer. **(C)** A representative single optical plane from a serial two-photon coronal section of the visual cortex showing detected tdTomato cells (red) and tissue autofluorescence (cyan) from a 1/100 dilution injection of AAV-PHP.eB-Cre. (Bar height - mean, error bars - SEM, N = 3 mice)

barcoded starter cells.

I performed tissue treatment in the same manner as with gene sequencing (Chapter 4 Methods 4.3.3), with some additional rabies-specific steps (Methods 5.3.3). In brief, rabies barcodes were detected using a padlock probe that is designed to bind to regions

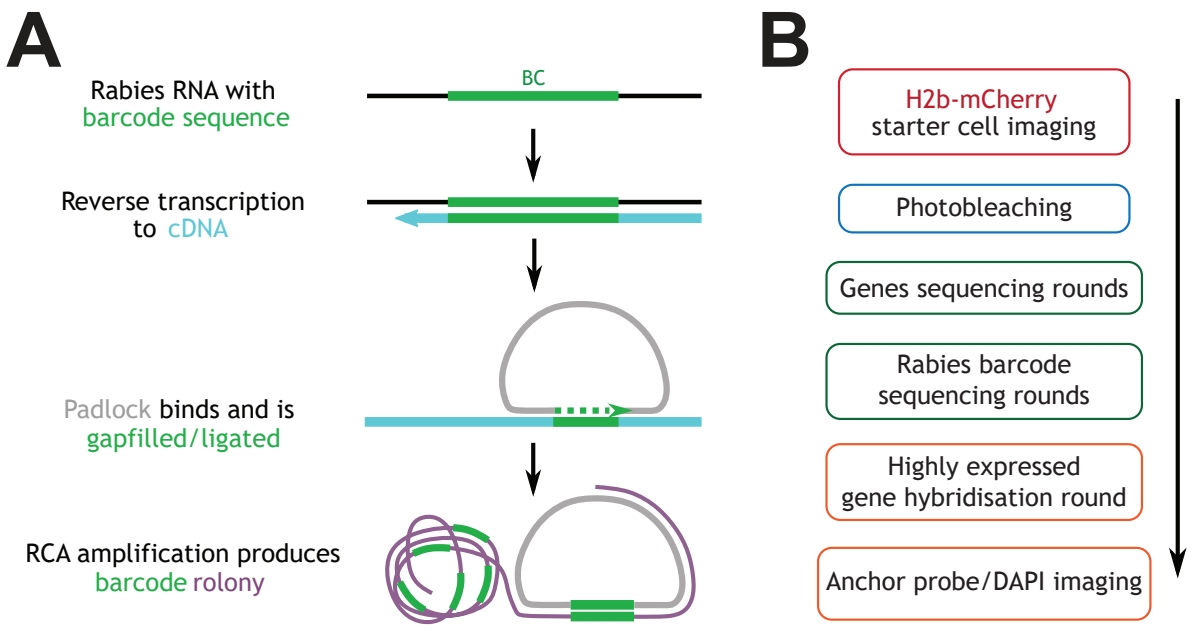


Figure 5.7: Barcoded rabies sequencing workflow. (A) A diagram of the rolling circle amplification of rabies barcode transcripts. **(B)** The imaging workflow for rabies sequencing experiments.

on either side of the 20 bp random barcode sequence. Rabies barcode viral transcripts are first reverse transcribed into cDNA (Figure 5.7 A). A single padlock probe is hybridised to the cDNA and Phusion polymerase is used to gap-fill the barcode sequence from the cDNA into the padlock. Then Ampligase is used to ligate the padlock into a piece of circular single-stranded DNA. The circularised padlock is then amplified using the highly processive phi29 polymerase to produce a DNA 'rolony' containing many copies of the rabies barcode sequence. These barcodes are sequenced in the same manner as gene barcodes (Chapter 4, Methods 4.3.3), in a separate set of imaging rounds that take place after the gene sequencing to prevent optical overcrowding of gene rolonies. The full rabies sequencing workflow begins with H2b-mCherry imaging to detect glycoprotein expressing starter cells, followed by photobleaching of mCherry and then imaging rounds of gene sequencing, then barcode sequencing. Finally, hybridisation probes for highly expressed genes are imaged, followed by imaging anchor probes and DAPI at the same time (Figure 5.7 B, Methods 4.3.3). These anchor probes bind to all of the genes/rabies rolonies and enable registration of all images to a common reference.

Unlike gene barcode sequences which are predetermined and can therefore be matched to a codebook of expected fluorescence across rounds, the rabies barcodes are random and must be base-called round by round accurately enough that the same barcode sequence can be detected and matched between cells in different tissue slices and flowcells. An imaging analysis pipeline was designed to perform this task (Methods 5.3.4) by detecting spots in the image, extracting spot fluorescence from these regions and then determining the highest likelihood base-call for each spot, per round (Figure 5.8 A). The initially predicted full-length barcode sequence is then used to generate an expected synthetic

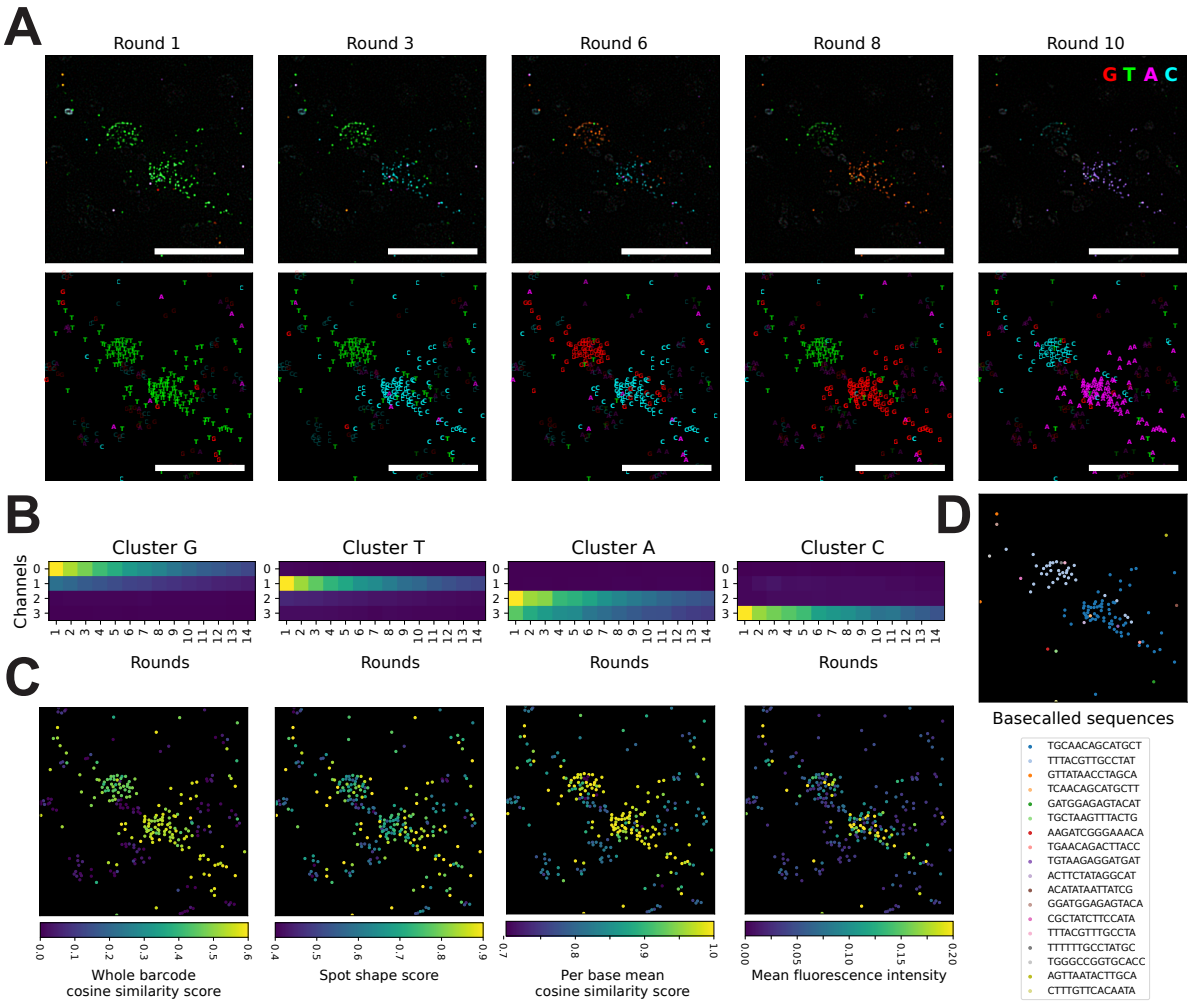


Figure 5.8: Rabies barcode rolonies can be detected in situ and base called. (A) In situ fluorescence signal from nucleotide conjugated dye sequencing of rabies barcodes. Spots are individual rabies barcode rolonies. Top row shows sequencing round 1,3,6,8, and 10. Bottom row shows base calls on each round of sequencing for each detected spot with opacity scaled to dot-product score of each spot (Scale bar 20 µm). **(B)** Bleedthrough matrices showing the mean rolony intensity values across channels and rounds for each of the four k-means separated clusters of nucleotide dyes. Values for each channel are normalised to the intensity of the first round of imaging. **(C)** Quality metric scores that are used to separate high-quality barcode calls from autofluorescence/background, including whole barcode cosine similarity, spot shape, per-base mean cosine similarity and mean fluorescence intensity. **(D)** Barcode sequences of rolonies that passed the quality metric thresholds of **(C)**.

fluorescence trace across channels and rounds, using a bleedthrough matrix constructed from the fluorescence values of the four dye-conjugated nucleotides from each round of sequencing (Figure 5.8 B). The observed fluorescence of each spot is compared to this synthetic trace to determine a cosine similarity score for the spot, which is used along with other metrics such as spot shape, per-base cosine similarity and mean fluorescence intensity (Figure 5.8 C) to filter out background fluorescence and noise from true barcode signals. The output is a series of barcode spot coordinates with their assigned sequence and metric scores (Figure 5.8 D).

Unlike gene barcodes, there is no reference sequence to try and match when base-calling

rabies barcode spots, so it is expected that individual base-called rabies barcodes will contain some errors. However, if certain conditions are met, these errors can be corrected. This first requires the true sequences within the experiment to be separated from each other by many base differences, which is expected given the diversity of the rabies libraries. It secondly requires the distribution of the number of errors per barcode to not overlap with the distribution of the base distances between the true barcodes. This overlap can be estimated using the structure of the data. In an experiment with many unique initial infections, rabies barcodes within different cells are expected to be different sequences, so measuring between cell barcode Hamming distances can give an estimate of the distance by which barcodes are separated. Rabies barcodes within the same cell are expected to mostly be the same sequence, from a single rabies infection producing multiple copies of the same barcode within a cell (although see Figure 5.11 B, E for rates of multiple rabies infections). Thus, if the same-cell base distances do not overlap with the between-cell base distances, we can be confident in error-correcting low-abundance barcodes with sequencing errors to their sequence neighbouring high-abundance "parent" barcodes.

After barcode assignment to cells (Figure 5.10), I calculated the distribution of hamming distances for barcodes that are found within a cell, and for barcodes found in different cells (Figure 5.9). Before error correction, 50.3% of the barcode pairs within cells were an exact match, with 27.1% of barcode pairs being mismatched by a single base, 9.8% of pairs mismatched by 2 bases and 1.9% of pairs mismatched by 3 bases. On the other hand, only 0.66% of sequences between cells had a minimum Hamming distance of 4 or less, meaning that the barcode error rate was sufficiently below that of the distance between barcodes. All of the base-called rabies barcode sequences from the whole experiment which passed quality control, were then error-corrected. Sequences were sorted by abundance and hamming distances were calculated between sequences. Lower abundance sequences which were within a hamming distance of 2 of a higher abundance sequence were error corrected to that sequence (Methods 5.3.4).

After Hamming distance 2 error correction, 86.6% of barcode pairs within cells were exact matches, with 0.006% with a mismatch of 1 base, 0.4% mismatched by 2 bases and 0.4% mismatched by 3 bases. The reduction in mismatches at three bases is likely to be due to the collapse of two separate mismatched barcodes onto the most abundant barcode in each cell. The remaining 12.6% of sequences with four or more mismatches after error correction begin to increase in abundance after a Hamming distance of 5 and most likely represent infection from multiple rabies viruses. 56,236 unique barcode sequences were detected in total before error correction. 4,771 barcodes remained afterwards. This represents only 0.45% of the possible 4^{10} 10 base pair barcodes. The distribution of Hamming distances of barcodes between cells on the other hand, did not change significantly before and after error correction, as the distribution begins to rise in abundance after a Hamming distance of 5, and so very few of these barcodes would have collapsed within a Hamming distance of 2.

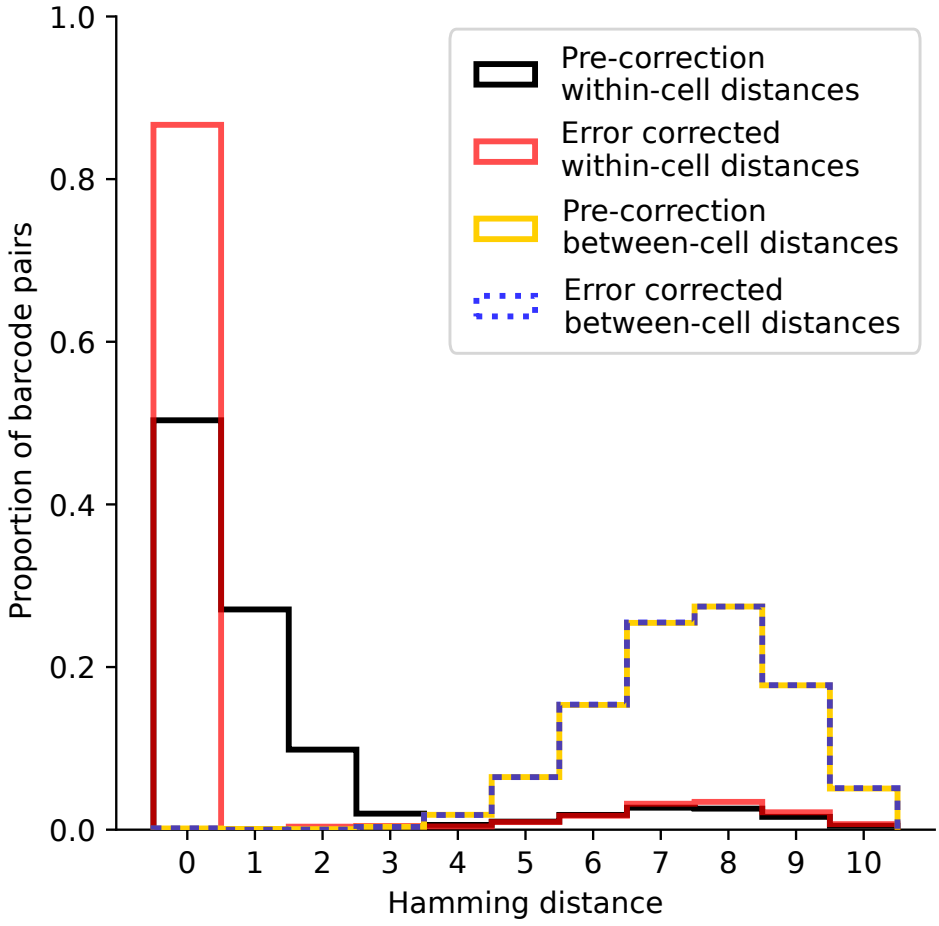


Figure 5.9: Hamming distance within cells is minimal compared to between cells, enabling error correction. Distribution of hamming distances between barcode sequences found within each cell, before (black) and after (red) error correction to the most abundant sequence within a hamming distance of 2, and the distribution of hamming distances of barcode sequences found between cells before (blue dashed) and after error correction to the most abundant sequence within a hamming distance of 2 (yellow).

Detected rabies barcode spots must be assigned to the correct cells so that the barcode information can be correctly linked to the transcriptomic information of each cell. This requires a marker of cell positions to assign the rabies barcode spots to spatially. As with the gene spots, I use a DAPI stain to define the location of the nuclei of individual cells (Figure 5.10 A). This is used alongside the anchor hybridisation probe (which labels all gene spots), to perform cell segmentation (Methods 4.3.4). The centroid locations of these segmented cell masks, along with the spatial location of the barcode spots and their sequences, are used to perform barcode assignment. The full details of the assignment algorithm are discussed in Methods 5.3.4. Briefly, the barcodes are assigned to proximal cell masks or to a 'background' assignment which accounts for sparse barcode spots that are found throughout the tissue far away from their soma of origin, from barcodes detected in neuronal processes. A Gaussian likelihood function defines the probability that a spot is associated with a cell centroid, along with a prior probability that accounts for the expected number of spots per mask, with a background component to account for the

aforementioned sparse barcode spots. Each set of barcode spots that corresponds to a unique barcode sequence is processed separately from other sequences, so the assignment of one sequence to a cell does not affect the assignment of others. The assignment is performed as an iterative process where each round of iteration attempts to reassign spots to cells to maximise the likelihood of observing this arrangement of spot assignments. An example of barcode assignment to the cell masks generated from the DAPI/anchor image in Figure 5.10 A is shown in Figure 5.10 B.

The H2b-mCherry fluorescence signal from the helper AAV1-FLEX-nuc-mCherry-N2cG must likewise be matched to the correct cell so that glycoprotein-positive starter cells can be identified, enabling the direction of connectivity to be determined for barcoded cells. The H2b-mCherry assignment is carried out by first defining masks from the mCherry fluorescence images and then performing quality control on the detected masks to generate a set of valid mCherry nuclei (Methods 5.3.4). These mCherry masks are then assigned to rabies-positive cells if they are found to overlap with rabies barcodes based on a set of spatial criteria (Methods 5.3.4). An example of four H2b-mCherry nuclei is shown in Figure 5.10 C. Of the four corresponding masks, three were determined not to be associated with any barcoded cells, whilst the fourth was assigned to a rabies-positive cell (Figure 5.10 D).

The rabies barcode assignment and mCherry assignments both rely upon the DAPI segmented cell masks to define cells. I observed that the cell segmentation resulted in the over-splitting of some cell masks into multiple small masks per cell and the under-splitting of others, resulting in merged cells. Also, as the DAPI stain is nuclear and rabies transcripts are produced in the cytoplasm of infected cells (Lahaye et al., 2009; Negri.A, 1903), the assignment of the rabies barcodes to these masks was not optimal. In recent attempts to find alternatives to DAPI-based cell segmentation, I have found that some primary antibodies will still recognise epitopes after the high-temperature steps of sequencing. Shown in Figure 5.10 E and F are antibody stains for NeuN and alpha smooth muscle actin (αSMA), which may act as stains for neuronal cell bodies in future experiments. Likewise, in Figure 5.10 E the stain for vimentin strongly labels blood vessels and could be used for image registration between tissue sections.

With barcodes and mCherry masks assigned to the cells, these barcoded neurons were grouped into putatively connected networks. mCherry-positive cells were defined as starter cells and mCherry-negative cells as presynaptic cells. I display the barcode-cell assignment data in Figure 5.11 as both the unfiltered data, and data obtained after applying filtering to retain only barcodes present in five or more presynaptic cells. This was performed to increase the overall confidence in the connectivity inferences drawn from the remaining barcoding data by removing small sets of barcoded cells which are numerically less likely to represent genuine transsynaptic spread.

Several notable features of this dataset can be observed from a summarisation of all of

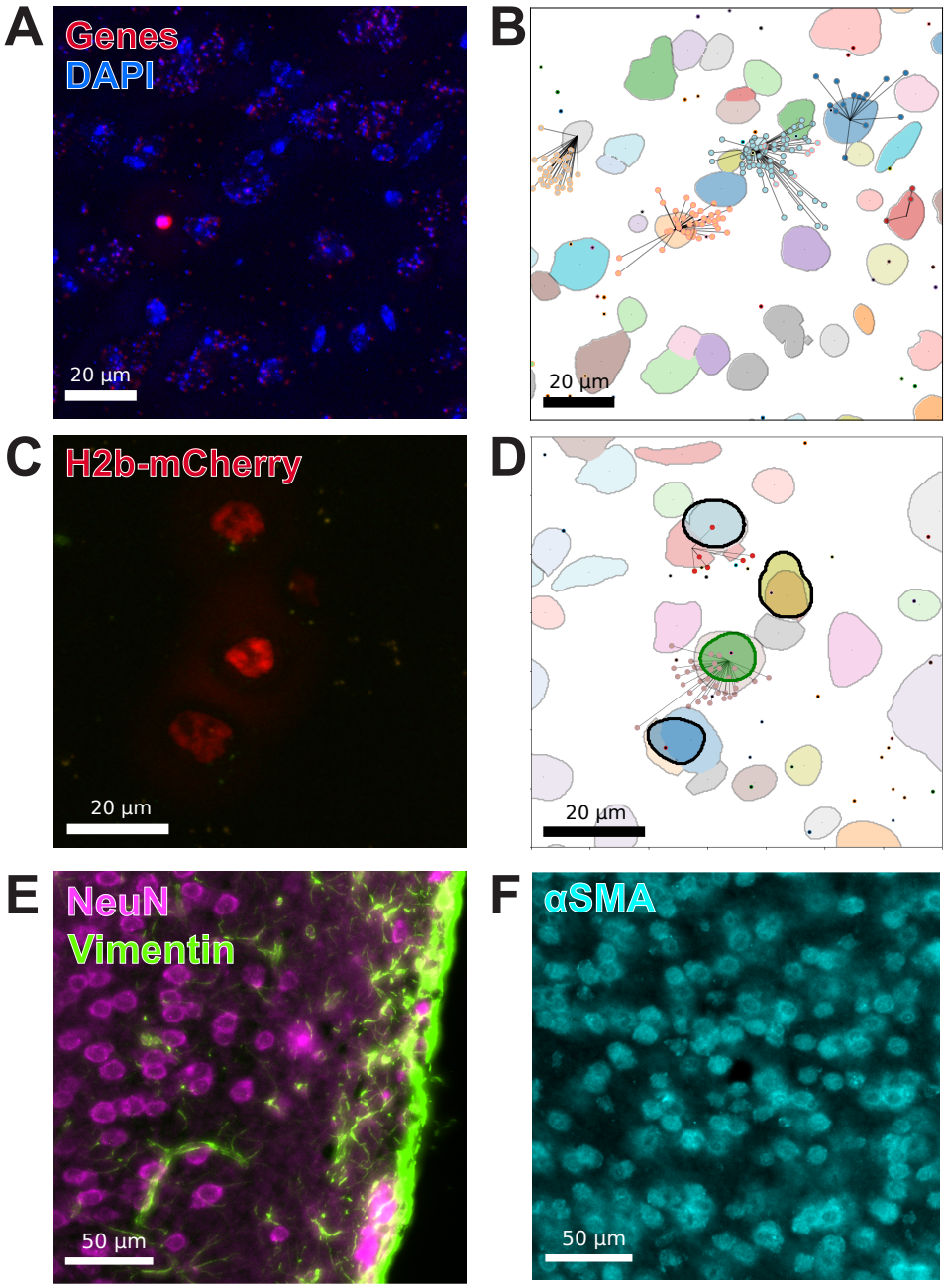


Figure 5.10: Rabies barcodes and mCherry nuclei assignment to cell masks. (A) DAPI nuclear stain (blue) and a fluorescent hybridisation probe against all gene padlocks (red) used for segmentation (see Methods). **(B)** Cell masks (colour-infilled contours) with overlaid rabies barcode spots (small circles). Barcode spots are coloured by their matching assigned cell mask with lines drawn to mask centroids. **(C)** Fluorescence image of H2b-mCherry from AAV-N2c-glycoprotein-N2A-H2b-mCherry infected cell (red) with autofluorescence in green. **(D)** A rabies-positive cell with rabies spots as described in **(B)**, and a spatially matching mCherry mask from **(C)** (green). Other mCherry masks from **(C)** are overlaid on rabies-negative cell masks (bold black lines). **(E)** Post-sequencing antibody stain for NeuN and Vimentin. **(F)** Post-sequencing antibody stain for αSMA.

the cell/barcode networks. For both presynaptic cells (Figure 5.11 A) and starter cells (Figure 5.11 D), the observed frequency of barcode rolonies per cell decreases exponentially, meaning that a large proportion of the detected cells have few rolonies per cell. The median rolony count per presynaptic cell is 9, with 13 median counts in starter cells. A

median of 4 presynaptic cells are found per barcode (12 after filtering), and the number of observed presynaptic cells per barcode also decreases exponentially (Figure 5.11 C). In presynaptic cells, the number of unique barcodes per cell is predominantly one (93.1% of cells) (Figure 5.11 B), whilst in starter cells, the proportion of multiply labelled cells is higher, at 28.4% (Figure 5.11 E). Some of these multiply barcoded starter cells are due to an initial multiplicity of EnvA rabies infection after viral injection. The remaining multiply barcoded starter cells are attributable to transsynaptic spread between starter cells. The number of barcodes found in more than one starter cell totals 108 barcodes, which are found across 273 starter cells (Figure 5.11 F). This results in 194 starter cells remaining which have barcodes that are not shared with any other starter.

Furthermore, out of the 2,260 unfiltered barcodes that were assigned to at least one cell, there are 1,689 unfiltered barcodes which are not found in any starter cell (Figure 5.11 F). These barcodes that are missing from starters may represent one of several scenarios.

1. These barcodes could belong to cells which were directly infected by contaminant rabies viruses which were not pseudotyped during viral packaging and retained their native glycoprotein coat, enabling them to infect TVA/glycoprotein-negative cells but then spread no further.

2. They could be a result of direct infection of cells which were positive for TVA but not glycoprotein, which likewise would result in no further spread of the rabies virus. This could be due to leaky expression from the TVA expression cassette.

3. They could result from an initial infection of a glycoprotein-positive cell, which resulted in successful infection of presynaptic cells after which the starter cell was lost due to cell death, physical damage to the tissue, inefficient detection of a weakly expressing starter cell's barcodes or H2b-mCherry signal

4. Sequencing errors in the starter cell/presynaptic cell could mean that the starter and presynaptic cells were not associated together correctly.

In the first two scenarios, with a sufficiently diverse barcode library, you would expect to find only one, or a small number of mCherry-negative cells which contain these barcodes as they would be individual infection events. In the third scenario, you would expect to find multiple cells sharing each barcode. In this case, the distribution of the number of cells sharing each barcode that is not found in starters would follow the same exponential dropoff as that observed for barcodes for which a starter was found (Figure 5.11 C). In the fourth scenario, sequencing errors could contribute to any part of the distribution of detected barcodes without starters.

To determine which of the proposed scenarios has contributed to the number of barcodes not found in starter cells, the distribution of the number of mCherry-negative cells found per barcode was plotted (Figure 5.12 A, B). 63.7% of barcodes not found in starters were

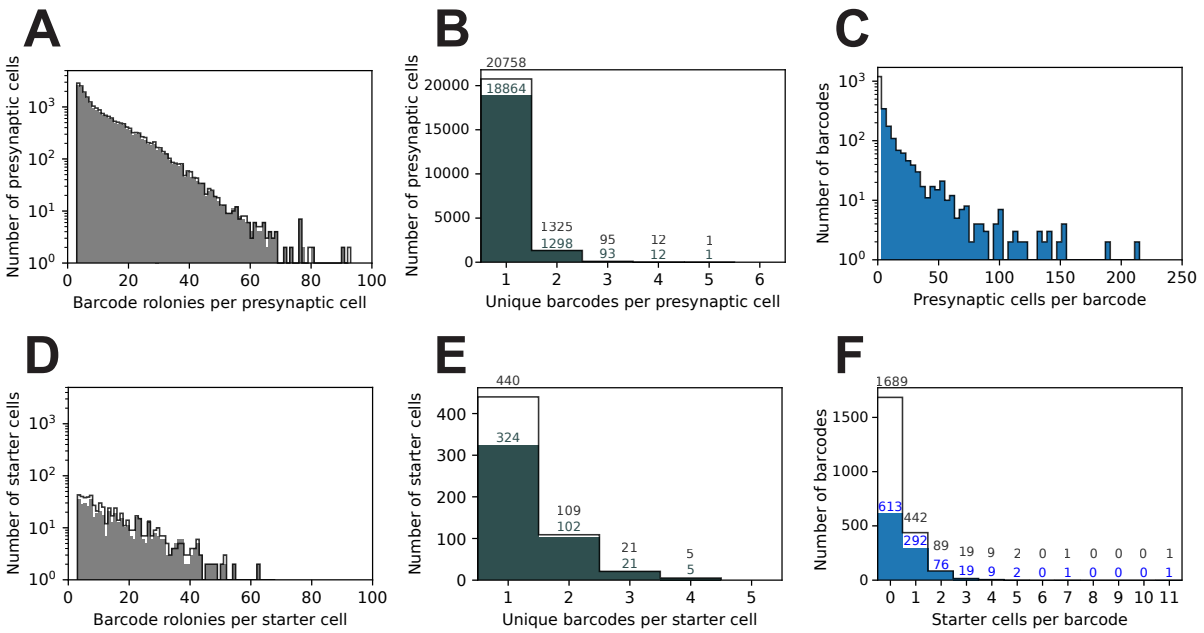


Figure 5.11: Barcode distribution among rabies-infected cells. **(A)** Distribution of the number of detected barcode rolony spots per presynaptic cell. **(B)** Distribution of the number of unique barcode sequences per presynaptic cell. **(C)** Distribution of the number of presynaptic cells found to share each unique barcode sequence. **(D)** Distribution of the number of detected barcode rolony spots per starter cell. **(E)** Distribution of the number of unique barcode sequences per starter cell. **(F)** Distribution of the number of starter cells found for each unique barcode sequence. Black lines on each plot represent unfiltered data, coloured bins represent data after filtering to only include barcodes found in 5 or more presynaptic cells (Unfiltered data: N = 22191 presynaptic cells, N = 575 starter cells. N= 2164 barcodes. Filtered data: N = 20268 presynaptic cells, N = 452 starter cells, N = 1174 barcodes).

found in less than 5 mCherry-negative cells. 31.1% of these were single-detected infection events which could have resulted from scenarios 1, 2 or 4. The tail of the remaining distribution included barcodes found in up to 121 mCherry-negative cells per barcode, representing cells likely to have come from scenario 3 or 4.

To determine the number of direct infection events that would be expected from the factors of TVA leak and glycoprotein-coated rabies contamination, control injections were made using the same injection conditions as the full barcoded rabies experiment. Injections of only barcoded rabies virus, with no helper AAVs expressing TVA, glycoprotein or AAV-PHP.eB-Cre were made into the visual cortex. This ensured that any resulting rabies infection must be due to glycoprotein-coated contaminant rabies, as without the helper viruses the EnvA-coated rabies should not be capable of infecting mouse neurons. After one week, the brains were fixed and the whole brain volume was imaged with the serial-sectioning two-photon microscope described previously. This resulted in the detection of 5 and 4 mCherry-positive rabies-infected cells in each brain respectively (Figure 5.13 A). The presence of infected cells demonstrate that whilst the 2 x 10^4 ratio of EnvA-coated rabies titre to glycoprotein-coated rabies titre found in Chapter 2 when producing the RV35 library is low (Figure 2.3), it is still high enough to permit small quantities of contaminant infection. However, this level of infection cannot account for a large

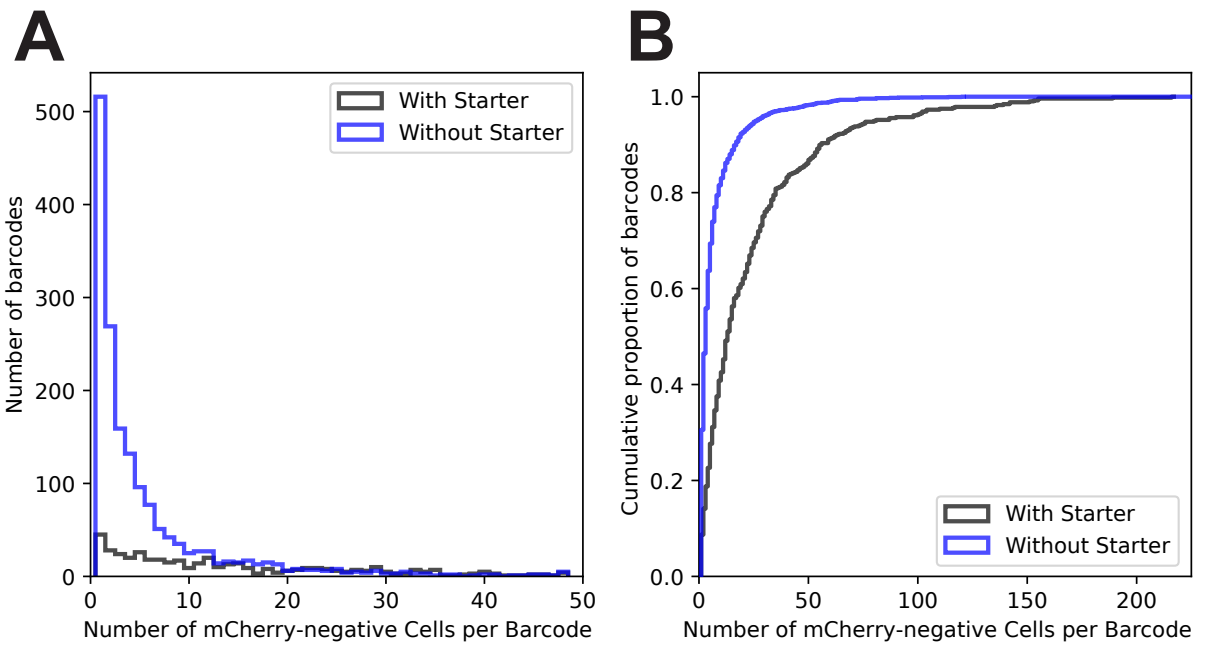


Figure 5.12: Barcodes not found in starter cells are predominantly found in few non-starter cells. (A) Distribution of the number of cells with a matching barcode for which a matching starter cell is present (black) or is not present (blue) and **(B)**, the cumulative frequency of this distribution (N = 1689 barcodes without a matching starter, N = 517 barcodes with a matching starter).

proportion of the 1689 unfiltered barcodes detected without starters.

A second set of control injections were made, first injecting only AAV-FLEX-nucEBFP2-TVA. One week later, as in the full barcoded rabies experiment, the rabies virus was injected at the same site. A week after that the brains were fixed and imaged (Figure 5.13 B). This set of injections, without the use of the AAV-PHP.eB-Cre virus allowed for the extent of the AAV-TVA "leak" to be measured. In a non-control experiment, using two viruses rather than a single construct to express both TVA and N2cG enables the dilution of each virus to be independently modified. This is particularly important as FLEX/Cre viral expression systems can be susceptible to 'leaky' expression in the absence of Cre caused by recombination events during plasmid production, or rare transcription events from the inverted transgene (Fischer et al., 2019). Whilst these AAV vectors have been designed to minimise leak through the use of the ATG-out system, where the Kozak sequence is placed outside of the inverted transgene (Wall et al., 2010), they are still susceptible to recombination events. As only a small amount of TVA expression is sufficient to enable pseudotyped rabies to infect cells (Lavin et al., 2020; Seidler et al., 2008), the dilution of the TVA construct is reduced whilst maintaining a higher titre of the glycoprotein AAV, as glycoprotein expression level affects the degree of transsynaptic spread from starter cells (Callaway and Luo, 2015; Lavin et al., 2020; Miyamichi et al., 2013). The same dilution of AAV-TVA was used here as would be in a non-control experiment. By omitting the glycoprotein AAV helper virus too, any rabies infects that were observed would have come from either direct glycoprotein-coated contamination events, or EnvA-

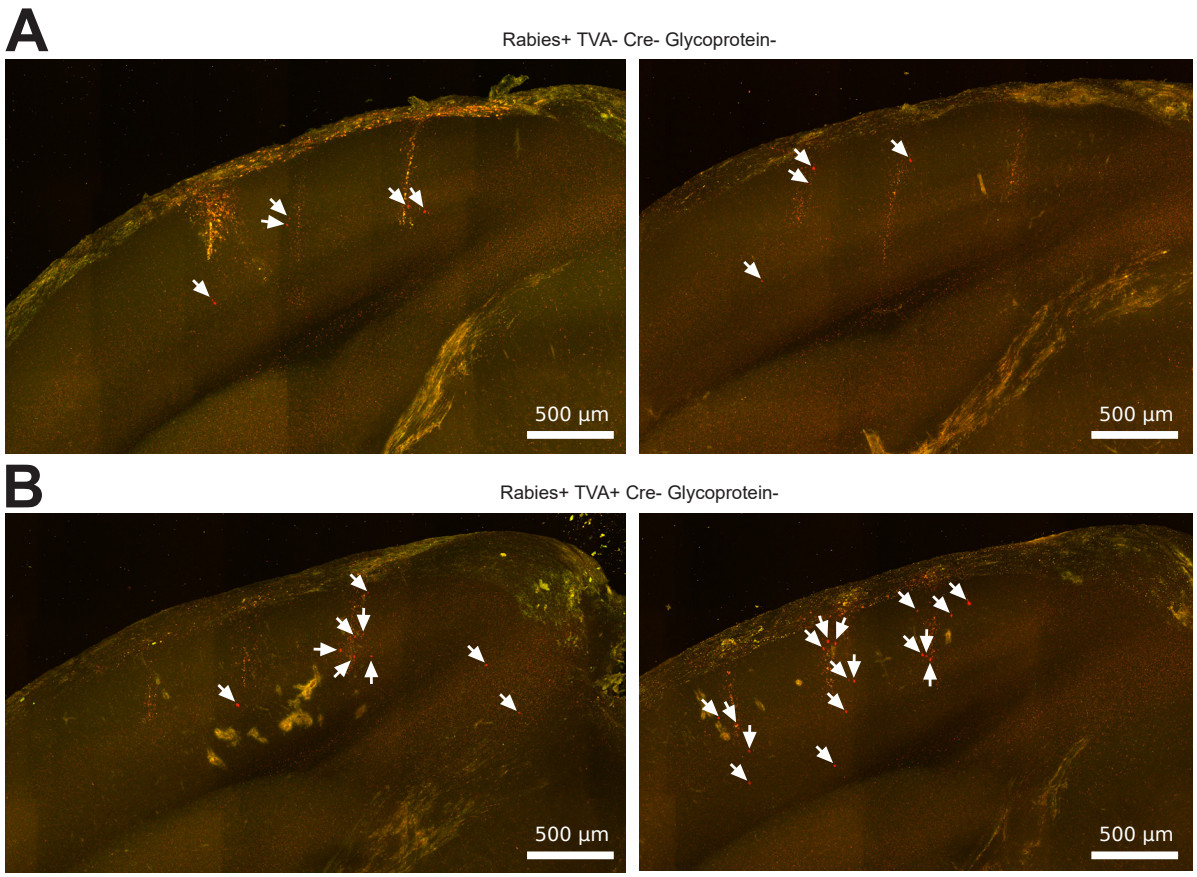


Figure 5.13: Control injections for AAV-TVA leak and glycoprotein-coated rabies contamination show minimal infection. **(A)** Serial two-photon maximum coronal z-projections displaying 1 mm thickness of tissue centred on the injection sites of barcoded rabies in two brains. No TVA/glycoprotein/Cre virus injections were performed. **(B)** Maximum z-projection through 1 mm of tissue from injection sites of both barcoded rabies and AAV-TVA in two brains. White arrows show the location of all detected rabies-infected cells.

coated rabies infecting TVA-positive cells from the leaky construct. A total of 9 and 17 mCherry-positive rabies-infected cells were detected, indicating that a small degree of glycoprotein-contamination / TVA-leak-induced direct infection events are liable to have occurred. Together, these control experiments demonstrate that the barcoded cells that were detected in the in situ sequencing experiment must have predominantly come from scenario 3 or 4 (initial infection of a starter, followed by loss of the starter, or failure to associate starters/presynaptic cells due to sequencing error).

To determine the degree to which the detection efficiency of the rabies barcodes was contributing to the number of undetected starter cells, I designed a set of positive control markers for rabies infection which could be detected at the same time as the rabies barcode padlock. The positive control markers are comprised of a set of 63 padlock probes designed to bind to the rabies viral transcripts for the N, M, P and L rabies genes. RCA was performed on tissue from a barcoded rabies injection site and fluorescent hybridisation probes against the pool of rabies gene padlocks and the rabies barcode padlock were imaged (Figure 5.14 C, D, E). For 15.77% of the rabies-positive cells detected by the rabies

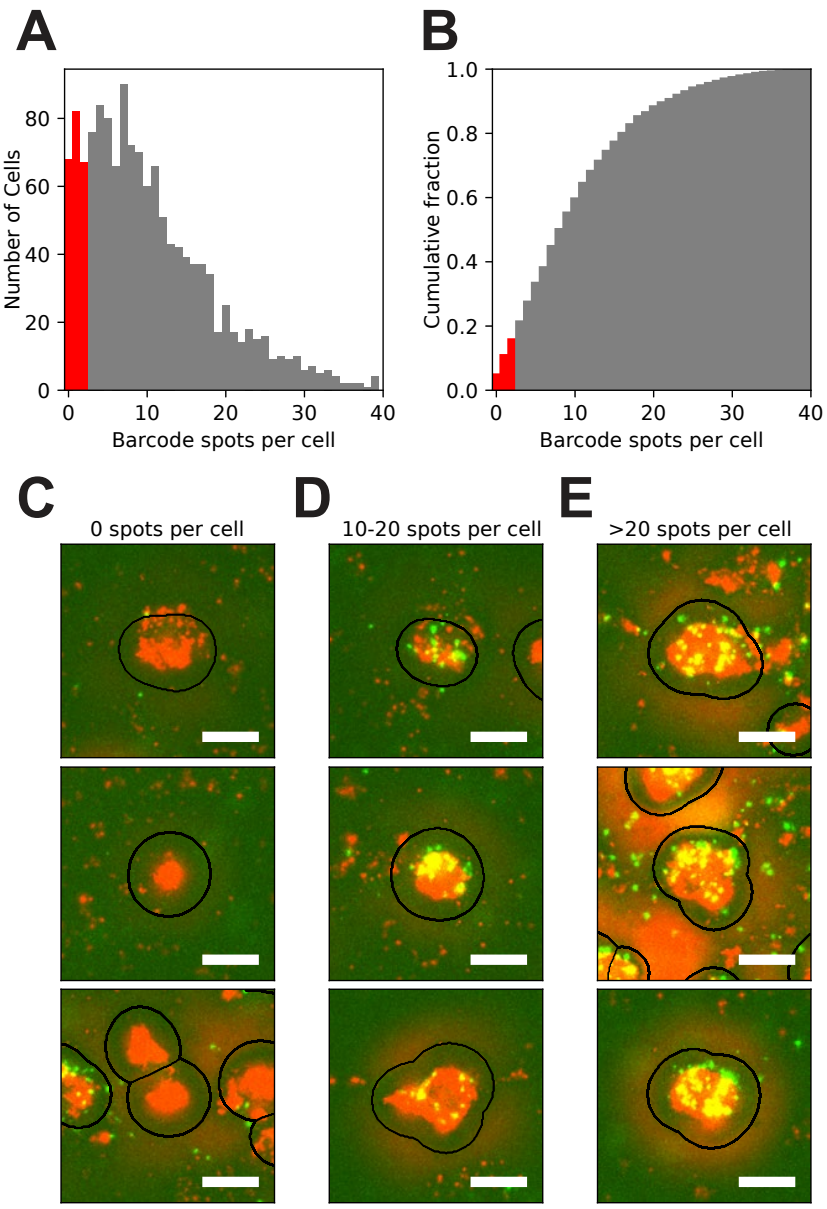


Figure 5.14: Limited rabies barcode detection efficiency from a single padlock. (A) Distribution of the number of detected barcode rolony spots per cell and **(B)** the cumulative fraction of the same distribution. Cells with <2 barcode spots are highlighted in red (N = 1376 total cells, 217 cells with <2 barcode spots, 15.77%). **(C, D, E)** Representative fluorescence images of randomly sampled cells with rabies barcode fluorescent hybridisation probe (green) and rabies N,P,M,L gene fluorescent hybridisation probes (red) with 5 µm expanded cell mask borders (black contour lines). Cells with **(C)** 0, **(D)** 10-20 or **(E)** >20 detected rabies barcode spots (scale bar 10 µm).

gene padlock pool, two or fewer barcode spots were detected. Our barcode assignment algorithm does not promote the assignment of barcode spots to a cell unless at least 3 rolonies of the same barcode are in proximity of the cell. This means that according to this control experiment, at least 15.77% of rabies barcoded cells are below the detection threshold with the current barcode detection efficiency. This applies to the detection of both starter cells and presynaptic cells. It does not rule out the other possibilities of scenarios 3 and 4 from also contributing to these missing starter cells.

5.4.3 Spatial analysis of barcoded cells

The 40 barcoded rabies tissue cryosections were registered to the CCFv3 Allen mouse brain atlas (Wang et al., 2020) (Chapter 4 Methods 4.3.4). This enabled me to transform the unrelated XY coordinates of individual tissue pieces into a single reference volume and analyse the rabies cell and barcode spot positions as a whole dataset, with anatomical positioning. Figure 5.15 A shows a coronal view of all of the 562,515 rabies spots which were assigned to cells, pseudocoloured by barcode sequence, and Figure 5.15 B shows all of the 575 detected starter cells, also pseudocoloured by barcode sequence. The 3D Allen coordinates of the isocortex were then transformed into a flatmap cortical representation to enable analysis of rabies infection with cortical depth (Chapter 4 Methods 4.3.4, Figure 5.15 C). The depth coordinates of all 281,247 uninfected (Figure 5.15 D) and 13,120 infected cells (Figure 5.15 E) from the primary visual cortex were used to determine the distribution of cell density. The density of infected cells was then normalised by the density of uninfected cells to give a ratio of infected/uninfected density (Figure 5.15 F). This ratio of infected cells is low in L1, with peaks in deep L2/3 and the boundary of L4/5, with another low density region on the boundary of L5/6a and a peak in deep L6a/L6b. These peaks correspond to layers that display more rabies infection than would be expected from the density of the cells alone. As predicted from the AAV-PHP.eB whole brain imaging data (Figure 5.6), starter cells are densest in L5, followed by L2/3, with low density in L6a/b and L1.

Figure 5.16 gives a subsample of the total data, as projected through the coronal plane. I combined the spatial information from each cell with their inferred connectivity using their barcode sequences. For barcoded connectivity analysis, I restrict the dataset to all 194 of the uniquely barcoded starter cells and the networks of corresponding putative presynaptic cells originating from them, of which a total of 5,785 were found. This highlights four of the 194 barcode networks, showing the extent of rabies spread from the injection sites that were tiled across V1.

From the 194 uniquely barcoded starter cells and their corresponding presynaptic cells, I took the spatial coordinates of all presynaptic and starter cell pairs which share a barcode and subtracted the starter coordinates from the presynaptic coordinates to centre all the starter cells at an origin point. From this origin, the distribution of all presynaptic distances to the starter cells could be measured in one reference frame. To create a null distribution of distances which retains the spatial structure of the original data whilst breaking the relationship between connected presynaptic cells and their starters, I shuffled the assignment of starter barcode sequences between the starter cells and presynaptic sequences between the presynaptic cells. Then I determined the distances between the cells that shared barcodes with this new randomised connectivity. Shuffling was performed 500 times to generate the null distribution for comparison. In the unshuffled data, a local peak in density was observed in the medio-lateral (ML) and dorso-ventral (DV) axes with

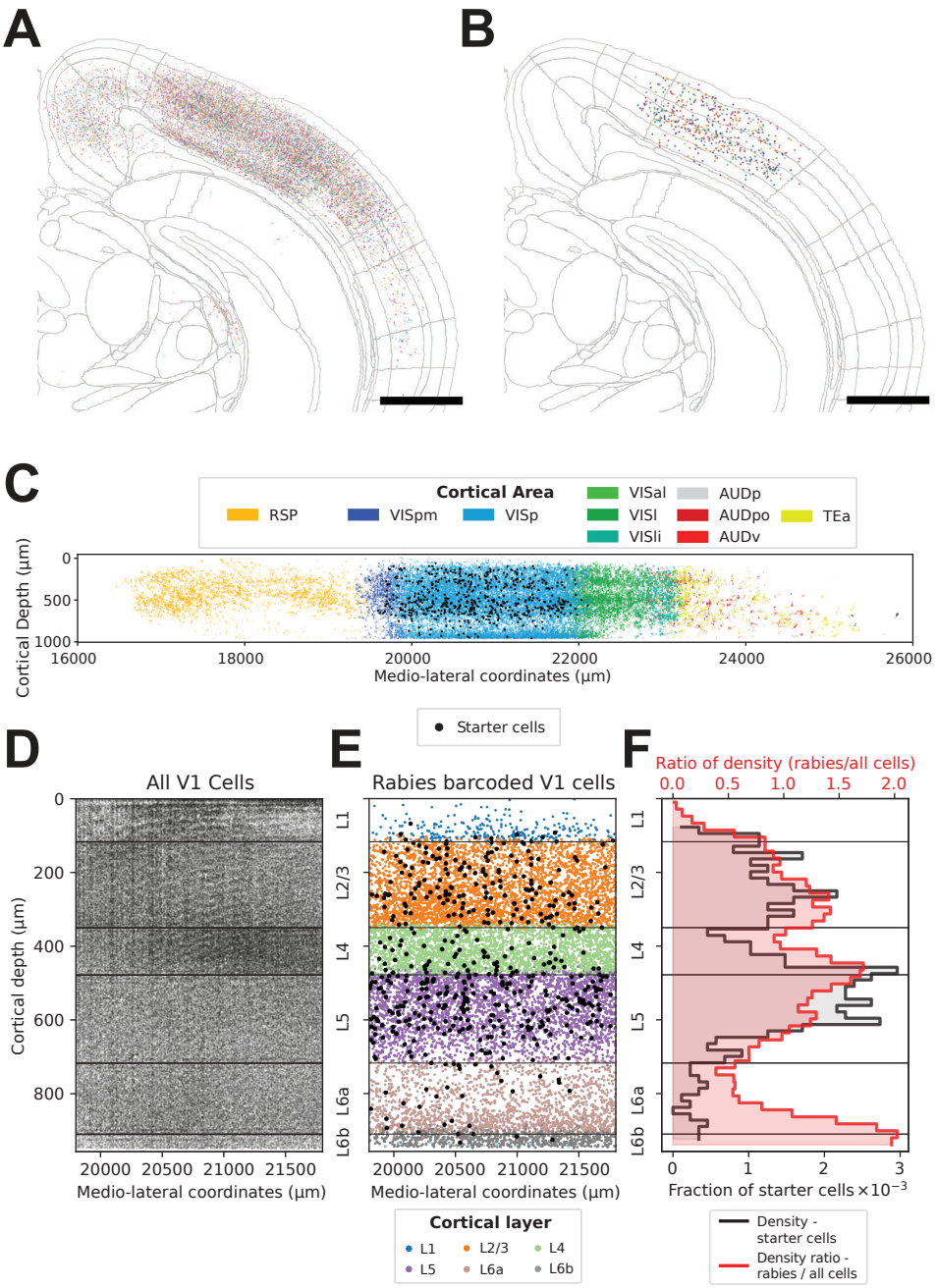


Figure 5.15: Barcoded rabies cells detected in situ show spatial density bias in cortical layers. **(A)** A coronal view of all detected rabies barcode spots, pseudocoloured by barcode sequence and mapped to CCFv3 brain coordinates (N = 562,515 barcode spots). **(B)** The same coronal view showing all detected starter cell locations, pseudocoloured by barcode sequence (N = 575 starter cells). **(C)** A mapping of rabies spot coordinates from **(A)** to a coronal view of a flattened representation of the isocortex. Rabies spots are coloured according to their cortical area of origin, starter cells are shown in black. **(D)** Coronal depth map of primary visual cortex with coordinates of all DAPI cell masks. **(E)** Depth map as in **(D)** with rabies-positive cell mask coordinates coloured by their registered cortical layer; starter cells are shown in black. **(F)** Distribution of starter cells (black line) across cortical depth, as a fraction per bin of total cells, with the ratio between rabies-positive cells and all V1 cells per bin in red. Scale bar = 1 mm

full width half maxima of 722.4 µm and 857.9 µm, respectively (Figure 5.17 A). The antero-posterior (AP) axis lies along the sectioning plane of the data and so only an 800

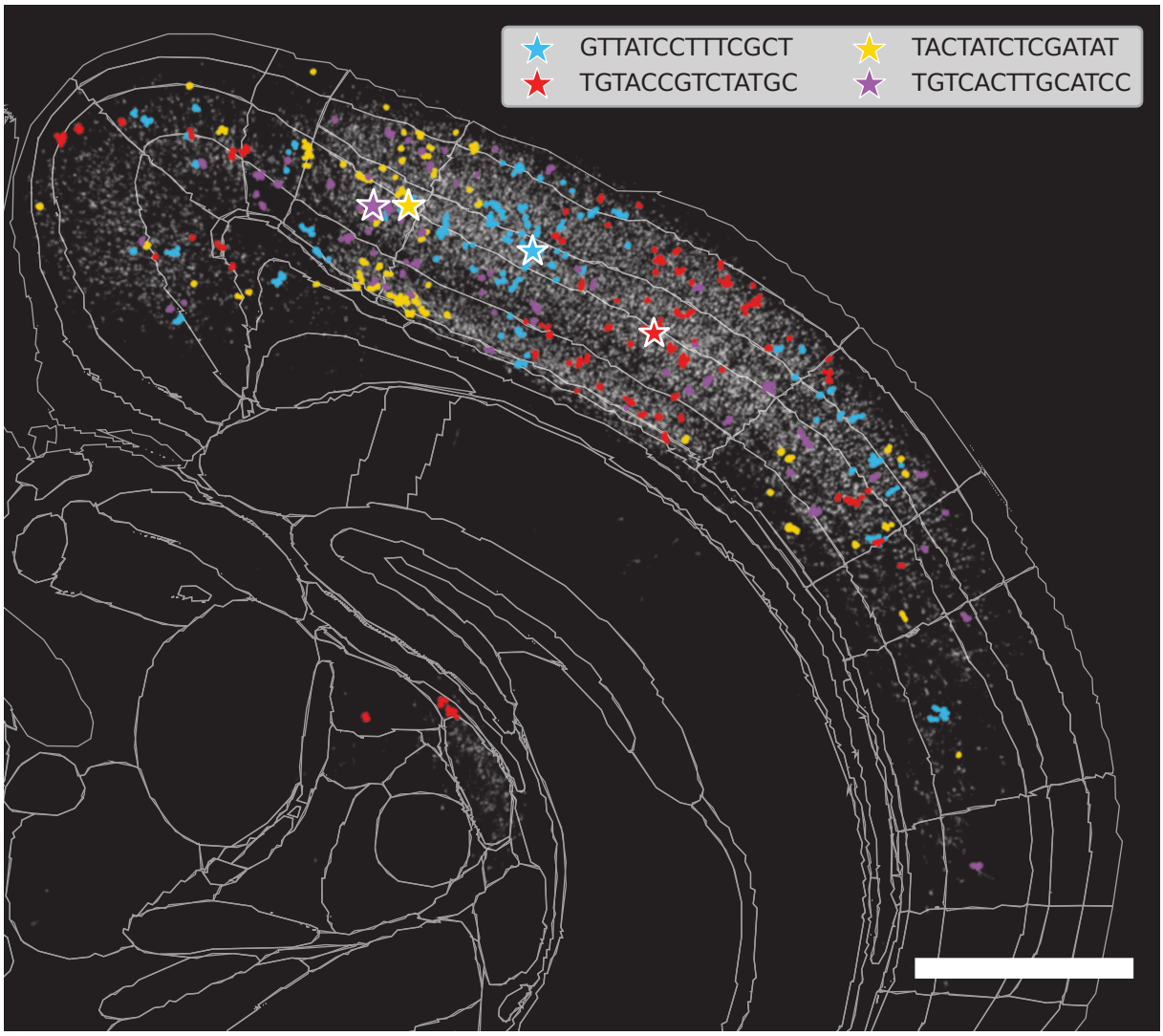


Figure 5.16: Barcode sequences matched between starter and presynaptic neurons. Coronal projection view of four example starter cells (coloured stars) and all of the presynaptic cells that match each of their respective barcode sequences (coloured circles). All other barcode sequences are plotted in grey. White outlines represent Allen CCFv3 area borders. Scale bar = 1 mm

µm window of the full cell distribution was sampled in this axis, giving a FWHM of 575.3 µm. The shuffled data still displays a peak in density with the outline of the cortex visibly present as spatial structure in the data is preserved, but the FWHM in the ML, DV and AP axes are broader at 2448.2 µm, 1294.3 µm and 689.0 µm respectively (Figure 5.17 B).

Comparing the 3D Euclidean distances between starters and presynaptic cells from the shuffled vs original barcode assignments enables comparison along a single axis. Notably, the median distance among shuffled-barcode pairs was 0.91 mm compared to 0.66 mm for shared-barcode pairs (two-sided Mann–Whitney U test, $p < 4.2 \times 10^{-56}$) (Figure 5.18 A). These findings suggest that cells sharing a barcode, and thus putatively connected, tend to reside closer together than those with different barcodes. These findings are consistent with previous evidence indicating that neuronal connectivity diminishes with increasing intersomatic distance, showing that putatively connected cells recapitulate expected spa-

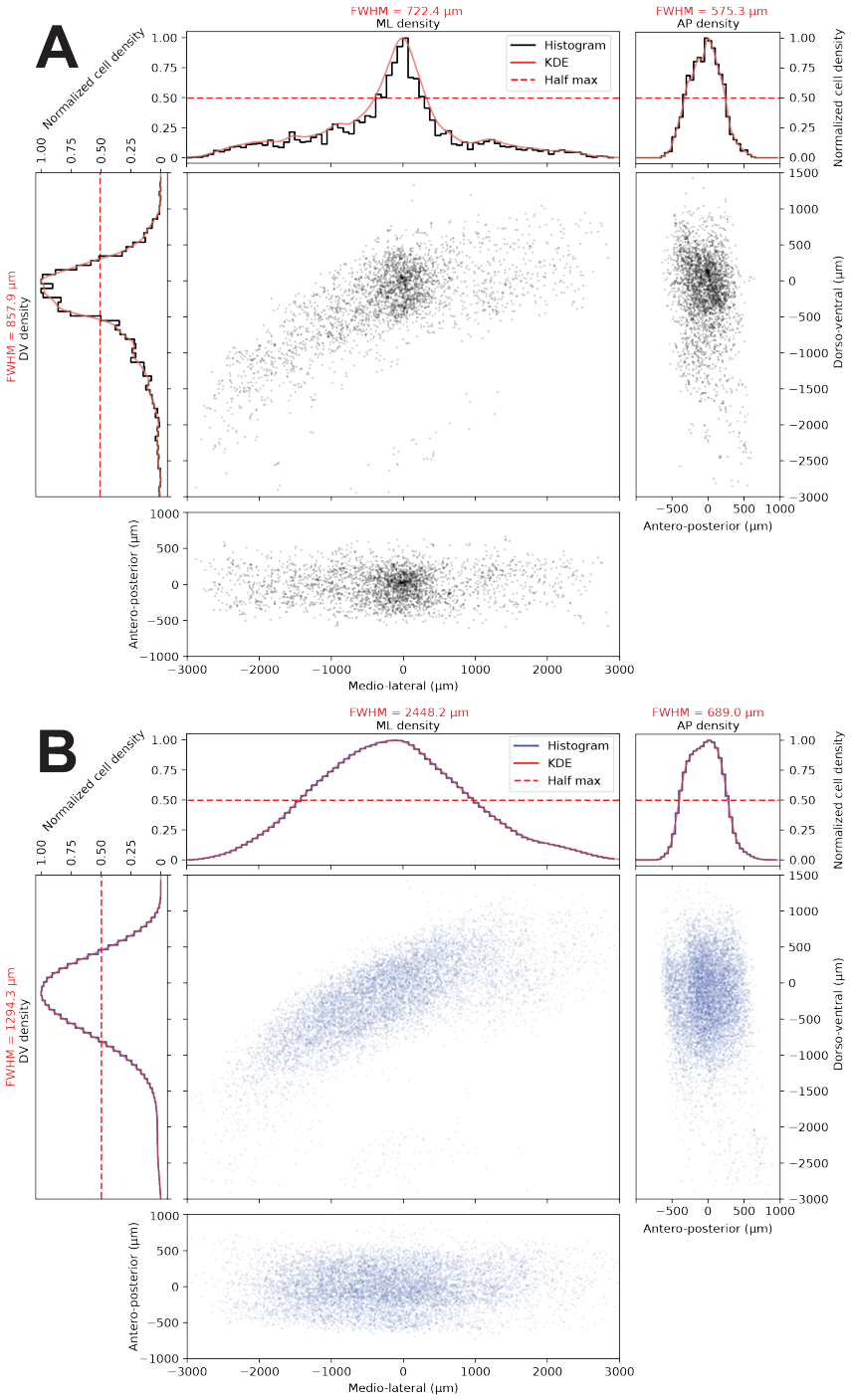


Figure 5.17: Distribution of distances from starter cells to barcode-sharing presynaptic cells and shuffled barcode control. (A, B) Relative coordinates of presynaptic cell locations from starter cells in CCFv3 coordinate space (presynaptic coordinate - starter coordinate) in coronal (centre), sagittal (right) and dorsal (bottom) views with corresponding normalised cell density distributions along the dorso-ventral (left), medio-lateral (top-left) and antero-posterior (top right) axis. **(A)** shows relative coordinates for all pairings of starter cells and presynaptic cells with matching barcode sequences. **(B)** shows a single shuffled null distribution, where barcodes between starters are randomly shuffled and barcodes between presynaptic cells are also randomly shuffled (N = 194 starter cells, N = 5,785 presynaptic cells)

tial structure (Campagnola et al., 2022; Holmgren et al., 2003; Perin et al., 2011; Song et al., 2005).

I carried out a similar comparison between pairs of starter cells, rather than starter cells and presynaptic cells to see if the same effect held for shared barcode starter pairs, which may have originated from transsynaptic spread between starters. I determined the 3D Euclidean distances between starters that share the same barcode. I then compared this to the distances found from those same starter cells to all starter cells which they didn't share a barcode sequence with, to create a null distribution of unconnected starters. The barcode-sharing starter cells were significantly closer together than the non-barcode-sharing cells by a two-sided Mann–Whitney U test ($p < 5.2 \times 10^{-11}$), with a median distance of 0.47 mm for barcode-sharing starters compared to 0.62 mm for non-sharing starters (Figure 5.18 B). The distribution of barcode-sharing starters appeared bi-modal, with a peak at less than 0.25 mm. It is possible that due to cryosectioning, a cell may end up split between two adjacent tissue sections, with a barcoded cell being detected in both. These physical duplicates in the data should share the same barcode. I performed the same 3D distribution analysis between starters but discarded cell pairs in adjacent tissue sections and counted the frequency of adjacent section cells with and without a shared barcode. I found that the proportion of adjacent section starter cells that shared a barcode was higher than that of cells that didn't share a barcode at 19.0% vs. 10%, respectively (Figure 5.18 C). When these adjacent section starter cells were removed from the 3D distribution analysis, the peak below 0.25 mm was removed, but barcode-sharing starter cells were still significantly closer than non-shared starter cell pairs by a two-sided Mann–Whitney U test ($p < 2.8 \times 10^{-4}$), with a median distance between barcode sharing non-adjacent starters of 0.52 mm. Together, these results show that as with starter and presynaptic cell pairs that share barcodes, starter cell pairs that share barcodes are also found closer together than those that don't. This suggests that these cells share barcodes due to transsynaptic spread rather than independent infection with the same barcode.

5.4.4 Connectivity motifs between brain regions and cell types

The previous analyses have all focused on the relationship between cell-cell distance and connectivity alone. Next, I incorporated two other pieces of information that I captured in the same experiment with in situ sequencing - the transcriptomic cell types of the barcoded cells (which were classified in Chapter 4), and the Allen atlas annotated brain regions that each barcoded cell belongs to. Together, these enable the analysis of location and cell type-specific connectivity motifs.

When represented in UMAP space, rabies-infected cells co-clustered with the non-infected cells (Figure 5.19 A, B). Furthermore, the distribution of total gene spots per cell for uninfected cells and rabies-infected cells was not significantly different (two-sided Kolmogorov–Smirnov test, $D = 0.01$, $p = 0.09$) (Figure 5.19 C). This is consistent with observations from previous in situ sequencing (Zhang et al., 2024) and single-nucleus RNA sequencing of rabies viruses (Patiño et al., 2022) which have demonstrated that cell

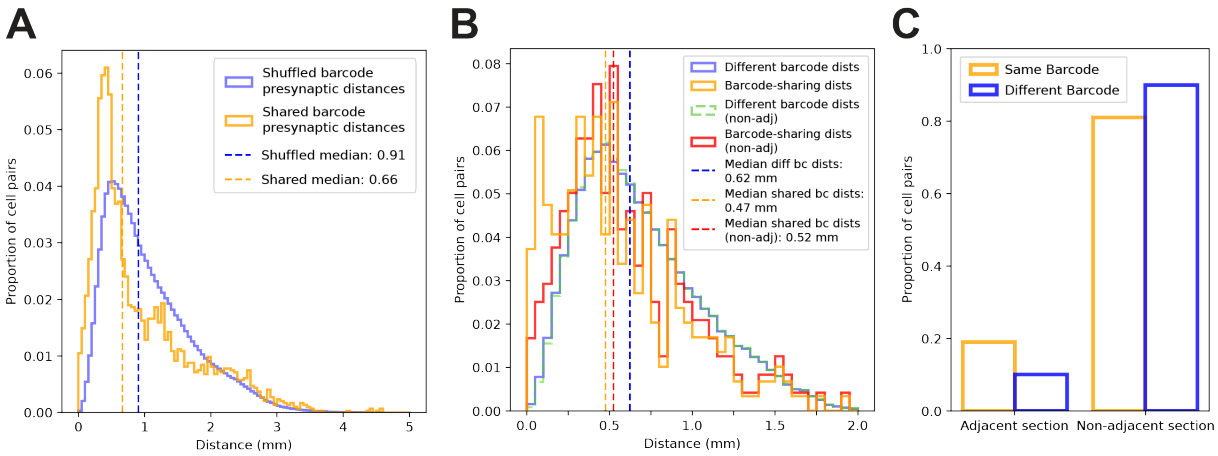


Figure 5.18: Presynaptic cells and starter cells that share barcodes. (A) Distribution of 3D Euclidean distances from starter cells to presynaptic cells that share the same barcode (yellow), compared to the distribution of distances for 1000 shuffle iterations where barcodes between starters are randomly shuffled and barcodes between presynaptic cells are also randomly shuffled. **(B)** Distribution of 3D Euclidean distances between starter cells which share the same barcode, including cells from adjacent sections (yellow histogram) or discarding them (red histogram), compared to the distribution of distances from barcode-sharing starters to all non barcode-sharing starter including cells from adjacent sections (blue histogram) or discarding them (green dashed histogram). Median values of each distribution are shown as vertical dashed lines. **(C)** The proportion of barcoded starter pairs with the same barcode (yellow) or different barcodes (blue), which are found in adjacent/non-adjacent sections.

type classification is possible in rabies-infected cells. The infected cells were distributed among all excitatory and inhibitory cell classes with similar proportions to the uninfected cells, although more L4, L4/5 and L5PT cells were observed (Figure 5.19 A, D), in line with the spatial distributions observed in Figure 5.16. Thus, no overt viral tropism is observable in the data.

Putatively connected cells can be grouped by brain region, cortical layer or cell type to find patterns of connectivity between different regions and types of cells. I first filtered my data to remove any starter cells that shared any barcodes with another starter cell. I then further filtered the data to remove networks of starters and presynaptic cells which have a total of less than five presynaptic cells, to prevent small networks of spuriously assigned barcodes from adding noise to the connectivity data. The resulting dataset contained 5,785 presynaptic cells and 194 starter cells.

This dataset predominantly consists of starter cells found in the primary visual cortex, with presynaptic cells also concentrated in the primary visual cortex, as well as higher visual areas and other local cortical regions within 800 µm along the antero-posterior axis of the brain (Figure 5.20 A, right and bottom total presynaptic and starter counts per area). As discussed in Section 5.4.3, starter cells from L5 and L2/3 are the most abundant. Outside of the primary visual cortex I have also sampled starter cells in the posteromedial visual area. Presynaptic cells are predominantly from L2/3 and L5, followed by L4 and L6. There are also significant presynaptic populations from the lateral

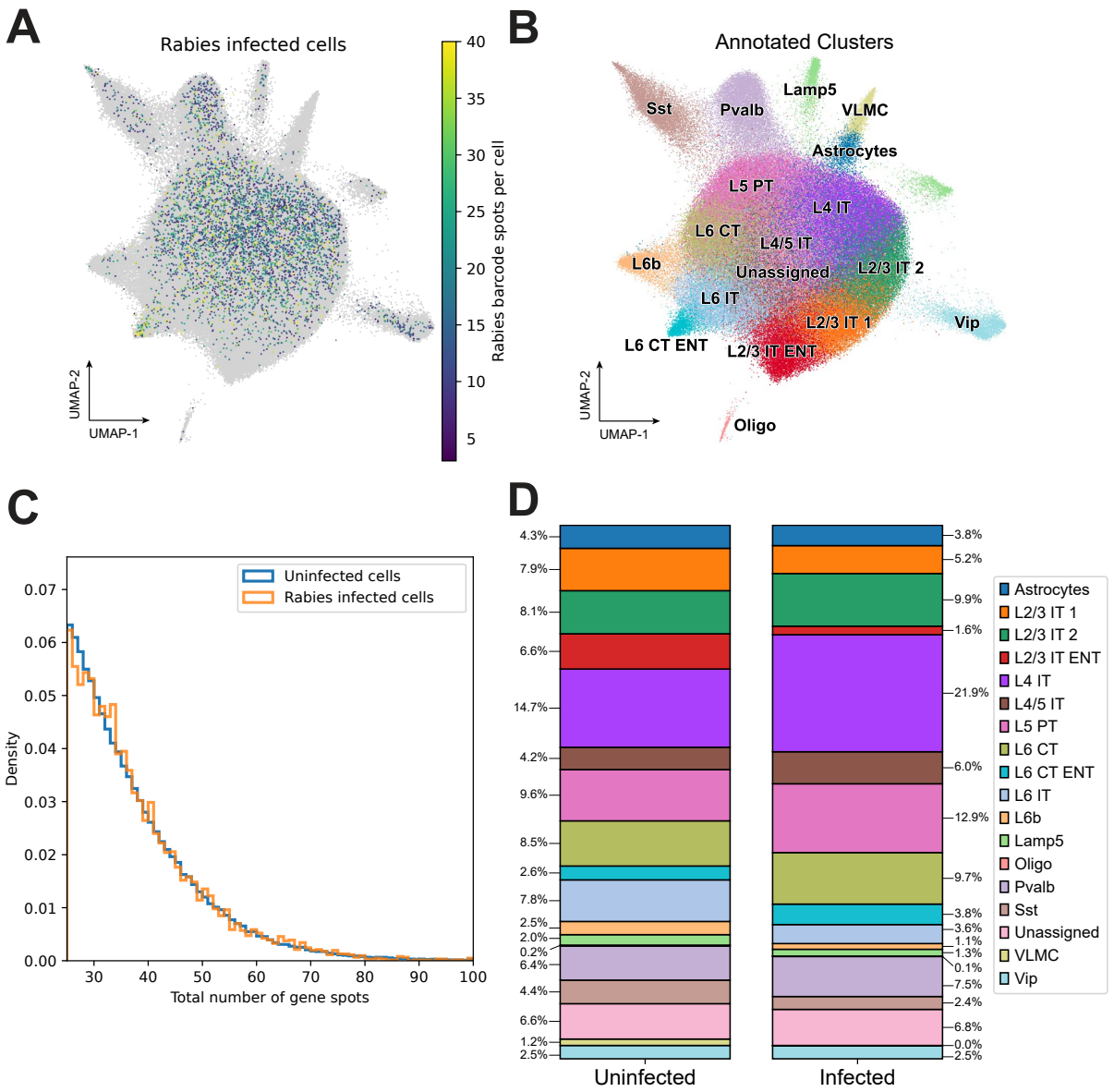


Figure 5.19: Rabies-infected neurons cluster with uninfected neurons and do not exhibit reduced gene counts. **(A)** UMAP representation of cortical cells clustered by detected gene probes. Grey spots represent uninfected cells, whilst coloured spots indicate the number of detected rabies barcode spots per cell. **(B)** The UMAP representation of cells in **(A)**, coloured by assigned cell type clusters with overlaid cluster labels. **(C)** Distribution of the number of gene spots detected per cell for uninfected cells (blue), rabies starter cells (green) and rabies presynaptic cells (yellow) (N = 452 starter cells, N = 20,268 presynaptic cells, N = 206,756 uninfected cells).

visual area, retrosplenial cortex and posteromedial visual area.

To visualise the relationship between presynaptic cell and starter cell locations, presynaptic cells were grouped by the location of their starter cell to make a matrix of area-to-area connectivity (Figure 5.20 A). As this matrix is dominated by the large differences in total presynaptic and starter counts between different areas, I normalised the number of presynaptic counts in each square of the matrix by both the total number of starter cells and total presynaptic cells corresponding to that presynaptic/starter location (Figure 5.20 B).

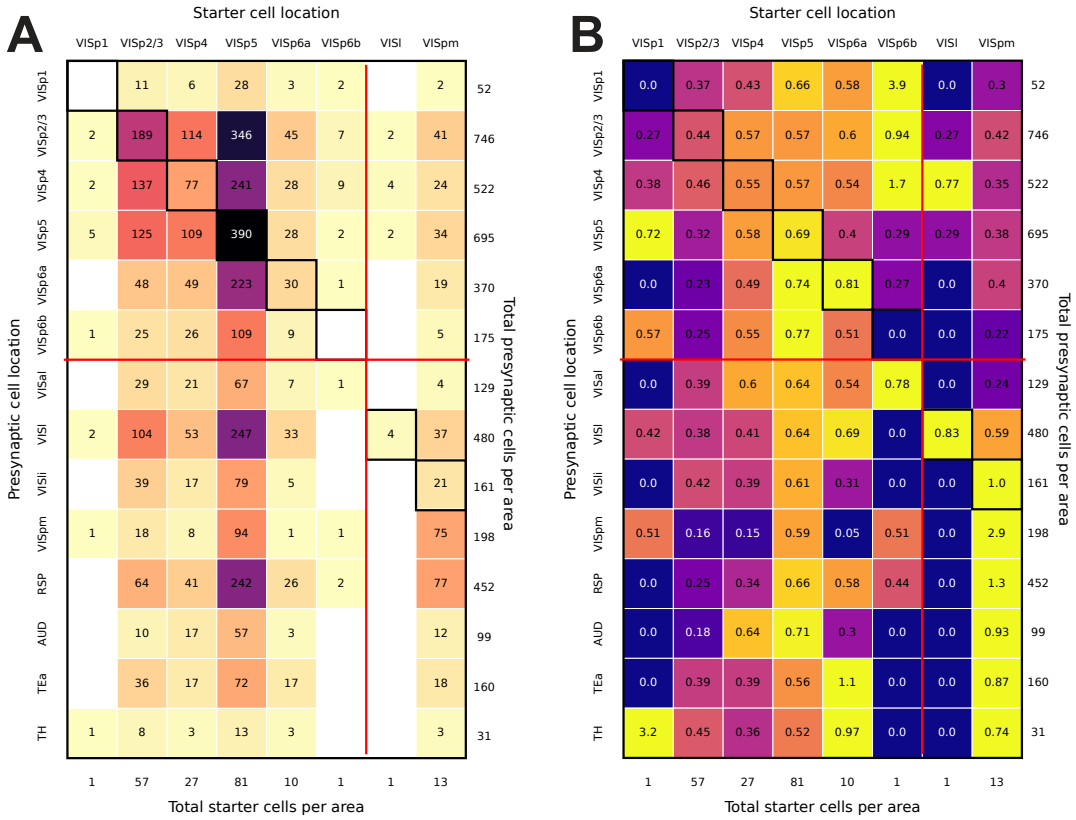


Figure 5.20: Connectivity matrix of starter and presynaptic cells by cortical area. **(A)** Connectivity matrix showing presynaptic cell counts grouped by presynaptic location and location of each presynaptic neuron's matched starter neuron, shaded by presynaptic cell count per area combination (count in grid), with the total presynaptic cells per area on the right of the matrix and total starter cells per area on the bottom. Red lines denote split between primary visual cortex regions and other cortical areas. **(B)** Normalised version of **(A)** where each cell value is normalised by both total starter cell count and total presynaptic cell count for that area, then scaled by a factor of 100 for percentage representation (N = 194 starter cells, N = 5,785 presynaptic cells).

Within the primary visual cortex, I observed that L2/3, 4 and 6 avoid input to L1, whilst L5 shows input to L1 and to itself. L2/3 cells connect to other cells from the same layer and input onto L4 and L5, whilst L4 shows input to L2/3, L5 and itself. L6a predominantly projects onto layers 5 and 6. These all recapitulate findings from previous AAV projection data (Harris et al., 2019). I observe input onto VISpm from higher cortical areas, including strong input from RSP, VISl, VISli, as well as input from the primary visual cortex, as has been found previously with monosynaptic tracing (Yao, van Velthoven, et al., 2023).

However, some findings do not agree with existing literature on the connectivity of the visual circuit. L6a and L6b also show high levels of input from L2/3 and 4, and lower levels of connectivity from L5 when the opposite is expected from local monosynaptic tracing from in the visual cortex (Yao, van Velthoven, et al., 2023).

In order to incorporate cell-type data into the rabies connectivity data, further filtering was required to select barcoded neurons which had adequate gene counts for clustering. A

gene count total of 25 or more was used as a cutoff for cell type clustering in Chapter 4, as below this value cells could not be adequately separated into clusters based on the limited transcript information (Methods 4.3.5). All cells that passed this threshold were clustered. After selecting all barcoded cells that were annotated with cell type clusters, the total number of presynaptic cells was 1792, with 84 starter cells. From this, the convergence of presynaptic cell location onto different starter cell types was observed (Figure 5.21 A, B). The most common starter cell types are L5 PT, L4 IT and L2/3 IT cells, in line with the laminar distribution of starter cells from Figure 5.20. Parvalbumin-positive interneurons are found to be the most abundant GABAergic starter cell type. This could be due to their higher relative abundance in the sampled tissue (Pvalb - 6.4%, Sst - 4.4%, Vip - 2.5%, Lamp5 - 2.0% of uninfected clustered cells), or due to the previously characterised tropism bias of AAV-PHP.eB for Parvalbumin interneurons (Brown et al., 2021; Jang et al., 2023).

L6 CT neurons received more long-range input from higher-order cortical areas than L6 IT neurons, with connections from VISl, VISpm, RSP, AUDp and TEa, consistent with single-cell-initiated monosynaptic tracing from whole-cell recorded L6 cortico-cortical and cortico-thalamic neurons (Vélez-Fort et al., 2014). Inhibitory starter neurons receive input from VISal, VISl, RSP, and VISli, recapitulating previous findings from cell-line-initiated monosynaptic tracing (Yao, Wang, et al., 2023).

Finally, I compared cell-type to cell-type input motifs (Figure 5.22 A, B). One cell-type circuit that has been particularly well-characterised is that of the Sst, Vip and Pvalb positive interneurons (Pfeffer et al., 2013). These interneurons exhibit connectivity via gap junction-mediated electrical coupling and through chemical synapses, but as rabies viruses are incapable of spreading via gap junctions (Tang et al., 1999; Ugolini, 2011), I compare my findings against previously determined synaptic connectivity. In this interneuron circuit, Sst cells inhibit Pvalb and Vip cells (Karnani et al., 2016; Ma et al., 2012; Pfeffer et al., 2013). Vip cells preferentially inhibit Sst cells with some inhibition of Pvalb (Lee et al., 2013; Pfeffer et al., 2013; Pi et al., 2013) and Pvalb cells inhibit mostly other Pvalb cells and Vip cells (Galarreta and Hestrin, 2002; Pfeffer et al., 2013; Tamás et al., 2000). My own dataset contains few instances of captured interneuron-interneuron connectivity, compared to the number of observed excitatory/excitatory (E/E), I/E and E/I connections. My available observations recapitulate Sst input onto Pvalb and Vip cells, Pvalb input onto itself and Vip, and Vip input onto Sst and Pvalb. I also find two Sst-to-Sst connections. A recent multipatch recording study of the mouse visual cortex also observed Sst-to-Sst connectivity throughout layers (Campagnola et al., 2022), which were confirmed to be between Sst cells via morphological matching, intrinsic electrophysiological properties of the cell and Patch-seq transcriptomic analysis. This new evidence was contrary to previous findings that whilst Sst-Sst electrical coupling is prevalent, chemical synapses had not been observed (Hu et al., 2011; Karnani et al., 2016; Ma et al., 2012). I observed L4 / L4/5 recurrent connectivity and L5 PT recurrent connectivity

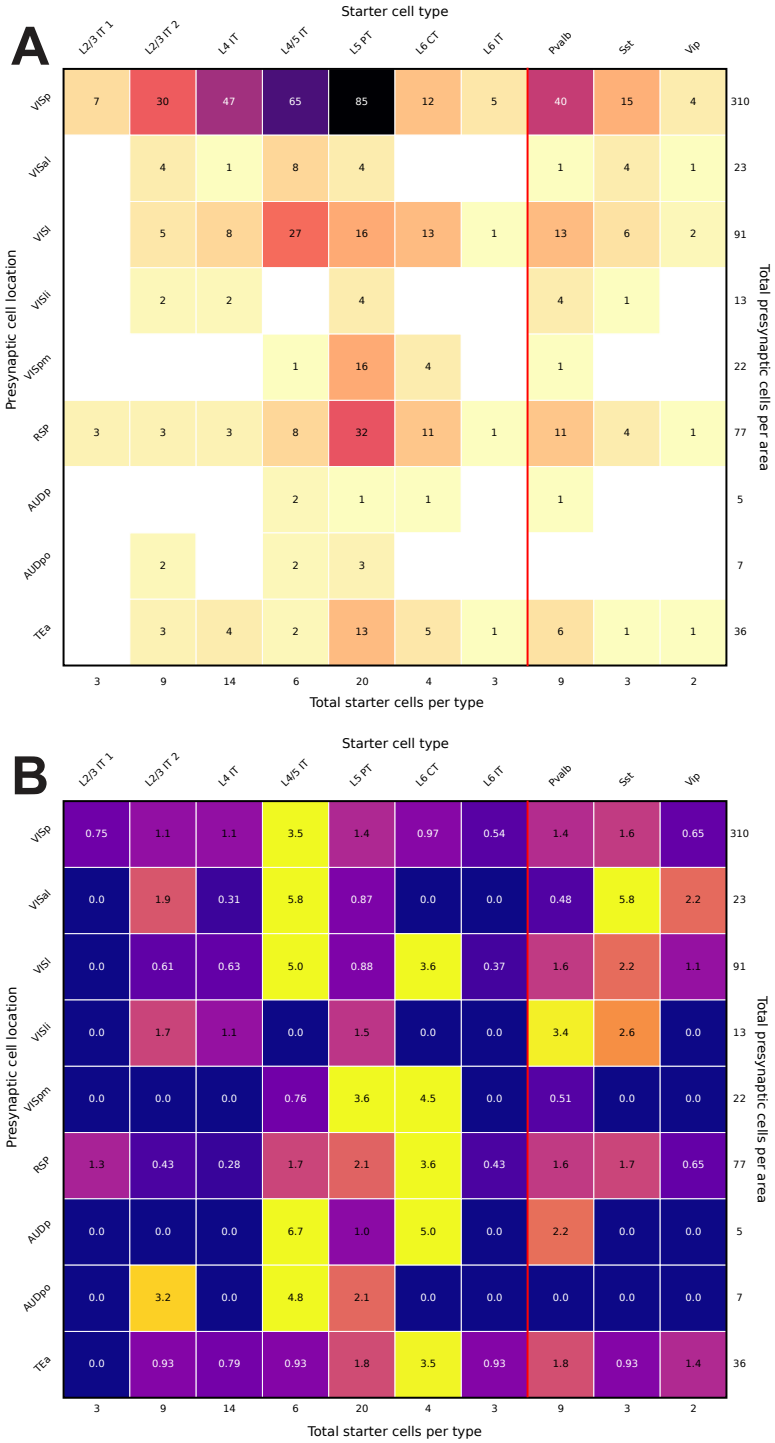


Figure 5.21: Connectivity matrix of presynaptic cell locations onto starter cell type. **(A)** Connectivity matrix showing presynaptic cell counts grouped by presynaptic location and cell type of each presynaptic neuron's matched starter neuron, shaded by presynaptic cell count per area/cell type combination (count in grid), with the total presynaptic cells per area on the right of the matrix and total starter cells per cell type on the bottom. Red line denotes split between glutamatergic and GABAergic starter cell types. **(B)** Normalised version of **(A)** where each cell value is normalised by both total starter cell count and total presynaptic cell count for that area, then scaled by a factor of 100 for percentage representation (N = 84 starter cells, N = 1,792 presynaptic cells after annotated cell type quality filtering).

but proportionally little L4 to L2/3 connectivity was observed compared to the canonical

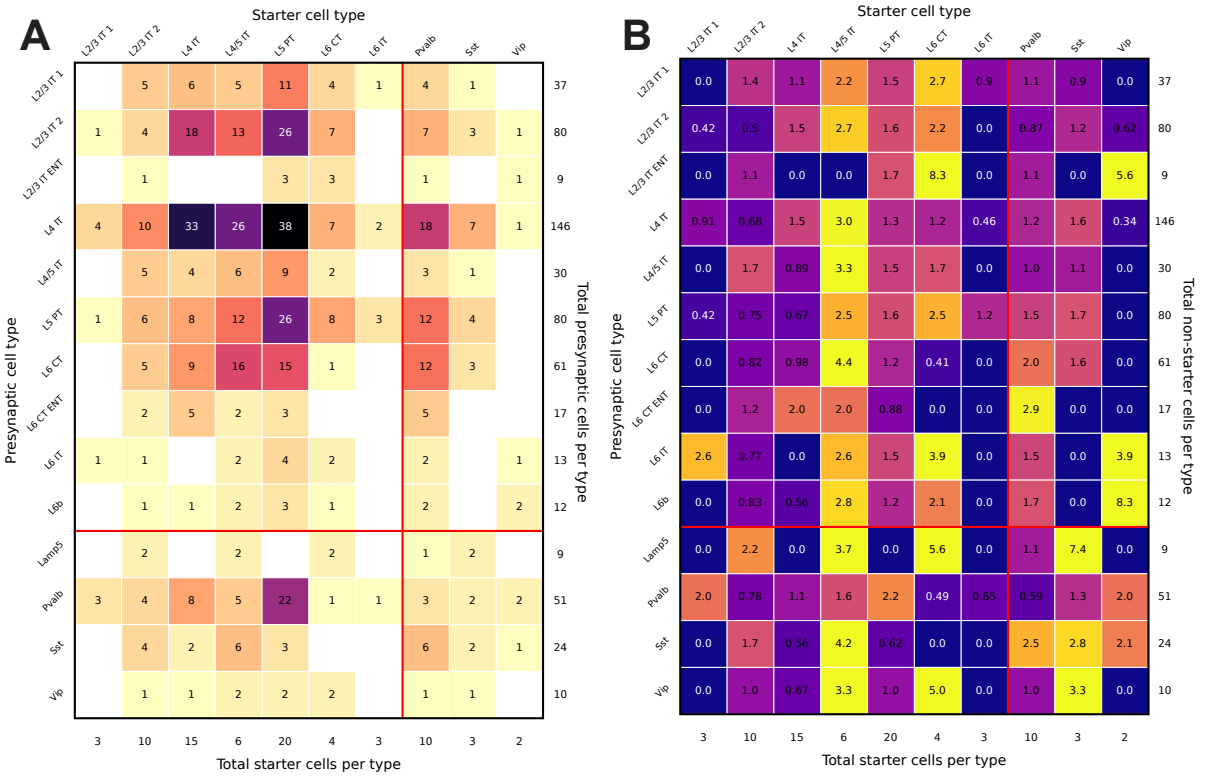


Figure 5.22: Connectivity matrix of presynaptic cell type onto starter cell types. **(A)** Connectivity matrix showing presynaptic cell counts grouped by presynaptic cell type and cell type of each presynaptic neuron's matched starter neuron, shaded by presynaptic cell count per cell type combination (count in grid), with the total presynaptic cells per area on the right of the matrix and total starter cells per cell type on the bottom. Red lines denote split between glutamatergic and GABAergic presynaptic and starter cell types. **(B)** Normalised version of **(A)** where each cell value is normalised by both total starter cell count and total presynaptic cell count for that area, then scaled by a factor of 100 for percentage representation (N = 84 starter cells, N = 1,792 presynaptic cells after annotated cell type quality filtering).

model of ascending visual information from the thalamus to L4 to L2/3 (Gilbert and Wiesel, 1979; Mangini and Pearlman, 1980), which has been observed with optogenetic and monosynaptic tracing measurements (Hage et al., 2022; Wertz et al., 2015). L2/3 to L5 connectivity is observed as expected (Harris et al., 2019), along with L5 input to L6 CT.

Overall, many of the findings from this dataset already largely agree with previously defined interareal connectivity, as well as interlaminar and cell-type specific connections found in the visual cortex. However, the data described in this preliminary barcoded connectomics analysis should not be taken as a resource for discovery or any inference of new connectivity rules, but rather as a proof of principle that demonstrates the scale and multi-omic capability that is afforded by this technique. Importantly, the analysis of this data has provided insight into the factors which can be improved for future barcoded connectomic experiments.

5.5 Discussion

By applying barcoded rabies virus libraries to the mouse visual cortex, I have demonstrated the capability to detect rabies barcodes from tissue and to use these barcodes to reconstruct the connectivity of thousands of cells from a single mouse at the resolution of individual neurons. A number of other recent studies have sought to use barcoded rabies viruses to determine neuronal connectivity (Saunders et al., 2021; Shin et al., 2024; Zhang et al., 2024, further discussion in Chapter 6.1.4). Each of these has highlighted a number of technical considerations that need to be overcome in order to use this new approach to produce unambiguous connectomic data. I will discuss the conclusions that can be drawn from the outcomes of my own experiments as to what has been improved over previous studies and what limiting factors remain to be addressed.

Ambiguity from shared barcodes among starter cells

In a monosynaptic tracing experiment, barcode-labelled rabies can spread only from cells expressing the rabies glycoprotein. As we are detecting connectivity from these glycoprotein-expressing starter cells to their presynaptic partners, it is the unique labelling of these starter cells which enables us to distinguish between all of the different convergent networks of connected neurons. As barcoded rabies viruses have the potential to spread between starter cells, this presents a problem for being able to unambiguously define which of the shared barcode starter cells a matching presynaptic cell is connected to. This problem was noted as a possibility in previous barcoded rabies studies. Saunders et al. attempted to reduce the likelihood of transsynaptic spread between starter cells in their in vitro cultured neuron experiments by sparsifying starter cell counts with a low titre AAV-Cre combined with high titre AAVs expressing rabies glycoprotein and TVA (Saunders et al., 2021). However, the extent of double-labelled starters that were observed in these experiments was not quantified and due to the necessity of dissociating and FACS sorting their samples, they were only able to recover around 10% of all infected neurons, which would have prevented accurate assessment of double labelling. Zhang et al. also noted the possibility of transsynaptic spread between starters and quantified this in their results (Zhang et al., 2024). They found that 27% of the 59 barcodes that they detected in starter cells were present in multiple starter cells. They estimated that most of these double-labelled starter cells were the result of spread between starters, given the diversity of their barcode library.

In my own experiments, I used data from whole brain cell counting to estimate a suitable dilution of AAV-PHP.eB-Cre for keeping starter density (and thus transsynaptic spread between starters) low while not compromising the total starter count per experiment more than necessary. I estimated this probability of transsynaptic spread between starters to be 8.1% for a density of 600 starter cells per mm^3 if all Cre-positive cells were infected with the

glycoprotein AAV vector. From the in situ cell/barcode data, I found 292 barcodes that uniquely labelled single starter cells and 107 barcodes that labelled more than one starter cell. These 107 barcodes that were found in more than one starter cell resulted in a total of 273 cell-barcode infections. These shared barcode starter infection events could have been caused by transsynaptic spread between starters, or they could have been caused by independent infection events from viral particles that shared the same barcode due to the skew of barcode abundance in the viral barcode library. To conservatively estimate the percentage of non-unique, independent infection events caused by my barcode viral library, I can use (Equation 2.2.2) from Chapter 2 to calculate the fraction of uniquely labelled cells for 715 individual infections (442 unique starter observations + 273 multiple starter observations), which gives a 3.0% non-unique infection rate. This would only explain 21 multiple labelling events, suggesting that the majority of double labelling observed in my experiments is due to transsynaptic spread rather than low barcode viral library diversity. Furthermore, I observed that the distance between starter cell pairs which shared the same barcode sequence is significantly shorter than the distance between starter pairs that do not share a barcode sequence, which is consistent with previously established evidence that neuronal connectivity diminishes with increasing intersomatic distance (Campagnola et al., 2022; Holmgren et al., 2003; Perin et al., 2011; Song et al., 2005).

When measuring the total number of detected mCherry-positive masks across the injection site of this experiment, there are a large number of uninfected mCherry-positive cells which would all increase the probability of spread from any infected starter cell to these uninfected cells. Counting infected and uninfected mCherry-positive cells together, there were 4122 total mCherry cells detected. This is more than double the expected starter density based on the mean density figures from initial PHP.eB titrations (Figure 5.2). This could be due to the inherent variability that we have observed from different batches of tail vein injections of AAV-PHP.eB which could be improved in future experiments by using a less technically challenging retroorbital sinus delivery route for the virus (Prabhakar et al., 2021; Tang et al., 2024).

Missing starter cells and barcode detection efficiency

I also observed in my data a large number of mCherry-negative barcoded cells, which contained barcode sequences not found in any starter cells. This constituted 1689 unique barcode sequences, of which 613 were present in 5 or more mCherry-negative cells. These barcoded cells without apparent starters could result from one of four scenarios. Firstly, the cells could originate from independent infection events of glycoprotein-coated rabies contamination from the viral library. Secondly, the cells due be the result of infection of non-glycoprotein-positive neurons due to leaky TVA expression. Both of these scenarios would allow the direct infection of cells without spread from glycoprotein-positive starter infections. Thirdly, the observation of infected cells without a starter could be due to

loss of the starter cells from cryosectioning damage, cell death or undetected/unassigned barcodes or H2b-mCherry from the starter cell. Finally, sequencing errors in the starter cell/presynaptic cell could mean that the starter and presynaptic cells were not associated together correctly.

From the control injections shown in Figure 5.14, the contribution of TVA leak or glycoprotein-coated rabies contamination is minimal and may explain only tens of missing starter events. The contribution of rabies cytotoxicity leading to starter cell death due to the prolonged time of infection within starter cells has been cited by Zhang et al. in their barcoded rabies study as their main source of missing starter cells (Zhang et al., 2024). Directly proving that starter cells have died due to the rabies infection is not possible with the data collected in this in situ sequencing experiment. It is possible that in future experiments, rabies-infected cell survival rates could be determined for barcoded rabies by observing fluorescent marker expression in-vivo with two-photon imaging over the 7-day time course of infection before the brain is extracted for sequencing. It must be noted that Zhang et al. used the SAD B19 strain of rabies virus for barcoding, whilst I chose to use the CVS-N2c strain of rabies virus for its reduced cytotoxicity and enhanced transsynaptic spread. If cytotoxicity due to rabies infection is proven to be a cause of starter cell loss, the time course of rabies virus infection may have to be optimised to reduce this loss of starter cells.

As part of the broader consideration of starter cell loss, it is possible that the starter cells can remain undetected due to tissue damage or low detection efficiency of barcoded rabies transcripts. One limitation of this dataset is that due to technical reasons, gene and barcode transcripts from a fifth flowcell that was at the anterior edge of the injection site were not efficiently detected, so the contribution of these putative starter cells could not be assessed. Within this missing flowcell, I conservatively estimated the total number of mCherry nuclei to be 100 from manual counting. Given the average rate of mCherry infection (575 rabies-positive mCherry cells from 4122 detected mCherry nuclei), this would mean the loss of 14 barcoded starter cells. I also observed from experiments that used a positive control set of rabies gene probes, that the single rabies barcode padlock was not detected efficiently enough to find barcode transcripts in all rabies-infected neurons. For 15.77% of the rabies-positive cells detected by the rabies gene padlock pool, two or fewer barcode spots were detected. Our barcode assignment algorithm does not promote the assignment of barcode spots to a cell unless at least 3 rolonies of the same barcode are in proximity of the cell. This gives a lower bound on the number of barcoded cells which may have been undetected in this current dataset. Finally, it is possible that some of the barcoded cells with no corresponding starter cell are, in fact, starter cells themselves. This could arise due to low expression levels of H2b-mCherry resulting in missed detection of glycoprotein-positive neurons. However, the AAV1-FLEX-nuc-mCherry-N2cG vector used for expressing the mCherry and glycoprotein in these experiments uses a bicistronic P2A peptide with the H2b-mCherry placed in the first position and the glycoprotein in

the second. 2A peptide constructs function by incorporating a linker sequence between two genes which causes the ribosome to skip the formation of a glycyl-prolyl peptide bond, effectively ”cleaving” the two nascent peptides (Donnelly et al., 2001). 2A peptide constructs have been demonstrated to have a much lower expression level of the protein in the second position due to ribosome fall-off after skipping (Liu et al., 2017), so very low levels of H2b-mCherry signal should mean that these cells express even lower levels of glycoprotein and would be incapable of facilitating rabies transsynaptic spread. Other barcoded rabies studies that have relied on scRNAseq rather than in situ sequencing have used glycoprotein or TVA transcripts as markers of starter neurons (Clark et al., 2021; Saunders et al., 2021; Shin et al., 2024). Zhang et al. did not use a nuclear-targeted fluorophore for their helper virus glycoprotein expression construct and, therefore, could not detect fluorescence from the cytoplasmic expression of their fluorescent marker which diffuses rapidly from fresh-frozen cryosectioned tissue upon thawing (Zhang et al., 2024). Instead, a set of padlock probes was designed against the rabies glycoprotein and used to determine starter cells. A direct comparison of the sensitivity of detecting H2b-mCherry or detecting the glycoprotein transcript with in situ sequencing should be made in the future to ensure the most sensitive method for detecting starter cells is implemented.

The effects of cell segmentation on barcode and gene assignment

In Chapter 4, I demonstrated the ability to simultaneously detect the gene transcripts of many target genes using in situ sequencing and showed that their spatial expression profiles match existing spatial patterns from in situ hybridisation data. Likewise in this chapter, I have shown that barcode spots can be reliably base-called with minimal errors and that the small number of errors which do occur can be corrected back to their most abundant “parent” sequences. The process of using DAPI for cell segmentation has resulted in a large proportion of low-quality cell masks in this dataset. This has manifested as the over-splitting of some cell masks into multiple small masks per cell and the under-splitting of others, resulting in merged cells. Furthermore, the transcripts of rabies viruses are produced in the cytoplasm of infected cells rather than the nuclei (Lahaye et al., 2009; Negri.A, 1903), and in situ detected gene transcripts are also found throughout neuronal cell bodies. This means that the nuclear staining of DAPI results in a cell mask which firstly does not capture the full extent of each cell’s transcripts, thus reducing the quality of data available for cell type clustering, and secondly, it is likely that in some cases rabies barcode transcripts corresponding to cells that are directly adjacent to those of a different cell type have been assigned to the incorrect cell due to these low-quality cell masks. I have since found that I can perform antibody staining, even after in situ sequencing tissue treatment and high-temperature sequencing rounds (Figure 5.10 E, F). This could enable improved cell segmentation through the use of one or more antibody stains, targeting cytoplasmic or cell membrane markers to define more inclusive cell boundaries. This could

also be used in conjunction with the rabies gene padlocks that I have designed, to provide an internal control for determining rabies barcode detection efficiency in future in situ experiments, as well as improving the cell masks of rabies-infected neurons.

Conclusion

In this chapter, I have described the application of a diverse barcoded rabies virus library to perform connectomic readout using in situ sequencing to detect rabies barcodes, whilst simultaneously integrating transcriptomic cell type information from rabies-infected neurons to produce a large, multi-modal proof of principle dataset. The use of AAV-PHP.eB-Cre has provided a way to control the density of starter neurons in a barcoded rabies experiment, and whilst DAPI cell segmentation has proven to be a limitation for the currently presented data, it allowed for the development of a working barcode-cell assignment algorithm and with the use of antibody staining for future experiments, the assignment of both gene transcripts and rabies barcodes could improve in accuracy. Barcode sequence analysis demonstrates that the error rate of rabies barcode base-calling is low enough for error correction, and analysis of the networks of putatively connected neurons demonstrated spatial clustering of cell pairs consistent with transsynaptic connectivity. The rabies-infected neurons in this data are transcriptomically clusterable into cell types and rabies infections were observed across a wide range of excitatory and inhibitory cortical neurons. The control experiments which were paired with the analysis of this large dataset of barcoded cells have addressed factors of rabies infection such as TVA leak and glycoprotein contamination and have drawn attention to the issue of barcode detection sensitivity as a necessary focus for future improvements to the technique.

Chapter 6

Discussion

6.1 Summary and general conclusions

In this thesis, I have described the development of a high throughput, distance-independent method for reconstructing neural circuits with single-cell precision. Crucially, I have demonstrated the feasibility of generating highly diverse libraries of barcoded rabies viruses which are capable of uniquely labelling thousands of neurons in a single experiment. By using in situ sequencing, I showed that these barcodes can be detected ex vivo from intact tissue sections and that the spatial information from in situ sequencing can be used to associate rabies barcodes, endogenous genes and helper virus markers to individual neurons. I applied these tools to construct a putative connectivity matrix that encodes both cell-type and brain region connectivity information at the single-cell level.

6.1.1 Producing high diversity barcoded rabies libraries

The viral libraries that I characterised in Chapter 2 represent some of the highest diversity rabies barcode libraries that have been described to date. Producing barcoded rabies libraries has only recently been attempted outside of the work described in this thesis, in part due to the limitations of low viral rescue efficiency imposed by the negative-strand genome of rabies viruses. The high level of diversity that I achieved in comparison to previously described libraries (Clark et al., 2021; Saunders et al., 2021; Zhang et al., 2024) is a result of optimising each stage of library production to retain barcodes through efficient plasmid transformation and viral rescue. This was followed by shortened viral packaging steps to minimise clonal expansion that can otherwise lead to a highly skewed barcode abundance distribution. Consistent with previous findings (Clark et al., 2021; Saunders et al., 2021), the largest loss of library diversity was observed during the viral rescue stage from plasmid DNA. Whilst the barcode libraries that I have generated do not reach the diversity achievable from other viral species such as adeno-associated

viruses, sindbis viruses and lentiviruses that can be rescued with high efficiency from plasmids (Adachi et al., 2014; Fennell et al., 2022; Kebschull, Garcia da Silva, Reid, et al., 2016), they provide diversity enough for uniquely labelling thousands of cells. Further improvements to rabies library diversity are also possible. My viral libraries were packaged on a small scale in comparison to other rabies packaging protocols (Osakada and Callaway, 2013; Saunders et al., 2021; Wickersham and Sullivan, 2015). I demonstrated an increase in library diversity with an increased scale of packaging, suggesting that large-scale libraries with higher diversity could be produced in the future, especially if combined with other improvements to viral rescue efficiency, such as using stable cell lines that express necessary packaging proteins (Shin et al., 2024; Sumser et al., 2022) and modifying the rabies packaging construct itself (Mori and Morimoto, 2014).

6.1.2 Increasing in situ sequencing throughput

Having produced sufficiently diverse viral libraries to perform connectomic experiments, I developed a method of barcode detection which could recover the highest percentage of rabies-infected cells from a brain as possible. As described in Chapter 5, the loss of barcoded cells and particularly the loss of starter cells that initiate the spread of barcode sequences to presynaptic cells, can be a confounding factor for determining connectivity in an unambiguous manner. The rate of cell retrieval after dissociation for single-cell RNA sequencing is too low for robust determination of synaptic connectivity from transsynaptic barcodes (Saunders et al., 2021). Meanwhile, in situ sequencing provided a means to detect transcripts from tissue without cell loss. BARseq, in particular, presented the unique ability to sequence arbitrary barcode sequences whilst resolving the position of each detected transcript with subcellular spatial resolution. This resolution was necessary for assigning barcodes accurately to cells in densely crowded tissue. Other spatial sequencing techniques such as Slide-SeqV2 (Stickels et al., 2021) and Visium (Maynard et al., 2021) can be used to detect barcodes (Lin et al., 2024), but the spatial resolution of these array based sequencing techniques cannot resolve individual cells. In Chapter 3, I described my efforts to increase the throughput of BARseq so that barcoded connectomics experiments that require the sequencing of large volumes of tissue for capturing all of the infected cells from a rabies injection, can be performed routinely. With sufficiently high experimental throughput, the ability to perform comparative connectomic studies becomes a possibility. By developing a custom imaging system and flowcell, I demonstrated that I could increase the imaging throughput of in situ sequencing over sixfold whilst increasing the amount of tissue that can be processed simultaneously. This equates to an imaging rate of 225mm^2 per hour. With the overhead of the manual sequencing chemistry steps, this results in an experiment length of around 2-3 weeks per mouse for completing gene and rabies barcode sequencing, assuming a similar volume to that described here in Chapter 5. I made progress towards fully automated sequencing by developing a microfluidics

system for handling sequencing reagents. However, this requires further optimisation to obtain consistent sequencing results. With an automated microfluidic setup, continuous imaging and chemistry steps would reduce the experiment length to 2.8 days. Automated sequencing would not only reduce the barrier of entry for using BARseq to perform rabies barcode readout but would also benefit all other applications of BARseq as a high throughput method for endogenous gene detection and the sequencing of other barcoded viruses.

6.1.3 Cell type and barcode assignment

In Chapters 4 and 5, I applied the high-throughput sequencing system that I had developed, in order to determine the connectivity of transcriptomically defined cell types. Clustering cells into assigned transcriptomic identities required the selection of a panel of differentially expressed marker genes. I demonstrated the in situ detection of the genes from this panel, each of which matched their expected spatial expression patterns. Using their spatial location, detected genes and rabies barcodes were assigned to individual nuclei-segmented cell masks. Transcriptomic cell types of neurons in the cortex follow laminar spatial patterns (Fang et al., 2022; Yao, van Velthoven, et al., 2023), and these patterns were used to assess the accuracy of cell type assignment. 74.1% of excitatory cells in the primary visual cortex were found in the layer expected of their assigned transcriptomic cell type. Both the assignment of these cells to transcriptomic clusters and the correct assignment of rabies barcodes to their cell of origin, rely on the accurate segmentation of neurons. The use of nuclei-based segmentation to assign transcripts to cells poses a problem for spatial transcriptomic techniques, as the nuclei alone do not capture the extent of transcripts in the cell body. Several approaches to assigning transcripts to cells in spatial transcriptomic data have emerged that use either prior knowledge from cell-type annotated scRNAseq datasets to probabilistically assign transcripts to cells (Petukhov et al., 2022; Qian et al., 2020), or deep-learning strategies based on spatial clustering of dense gene transcripts to segment cells (Chen et al., 2023; Chen, Xu, et al., 2024). Whilst these approaches may be applicable to the gene transcripts from BARseq data, improving the assignment of rabies barcodes to cells cannot rely on knowledge of prior expression patterns as with genes, and requires higher-quality cell masks and improved barcode detection efficiency to increase the accuracy of cell assignment. I have demonstrated that post-sequencing antibody staining can produce reference data for cell segmentation that extends beyond the nuclei of the neurons and I have also developed a sensitive method for detecting rabies-infected neurons by using a panel of probes that target rabies genes. Together, these will aid in improving the accuracy of cell segmentation which in turn will improve cell type and barcode assignment for future experiments.

6.1.4 Interpretation and comparison of connectomic information

The putative connectivity matrix that I characterised in Chapter 5 represents, to my knowledge, the largest reconstruction of synaptic connectivity using barcodes that has yet been described, in terms of both the volume of reconstructed tissue and the number of neuronal connections that were resolved. Recently, a number of other studies have described their own application of barcoded rabies viruses to study connectivity (Clark et al., 2021; Lin et al., 2024; Saunders et al., 2021; Shin et al., 2024; Zhang et al., 2024). Their methodological variations and the resultant connectomic information that they describe, serve to highlight some important considerations when performing barcoded connectomics experiments.

Multiplicity of infection

One outstanding question when infecting cells with barcoded rabies was whether starter cells could sustain multiple initial infections. A similar question could be asked of presynaptic cells - could they be infected by multiple barcoded viruses, either from the same starter or from different starter cells which shared a presynaptic partner? For many viral species, initial infection by one virus will exclude subsequent infection by a second virus if enough time elapses for the first virus to establish expression (Johnson and Spear, 1989; Kobiler et al., 2010; Meurens et al., 2004; Whitaker-Dowling et al., 1983). This has also been observed in rabies viruses, with rates of co-infection dropping, but not being fully abolished after a period of 6-12 hours between viral introductions (Ohara, 2009).

If rabies infection does produce a superinfection exclusion effect, it would be expected that the initial infections which occur immediately after viral injection would permit multiple rabies virions with different barcodes to infect starter cells, but that presynaptic infections from multiple starter cells would be more likely to be excluded as they would be subject to longer time interval between infection events due to variation in rates of amplification and transport to presynaptic neurons. Consistent with this time-dependent viral superinfection exclusion, I observed that both starter cells and presynaptic cells tended to contain a single viral barcode, but presynaptic cells contained fewer barcodes on average than starter cells (mean barcodes per starter - 1.28, mean barcodes per presynaptic 1.07, N = 575 starter cells, 22,191 presynaptic cells). This lower rate of multiple barcode infection in presynaptic cells could also be explained by lower probability of infection from starters than the initial high multiplicity of infection from EnvA injected rabies. The low rate of multiple infections was also observed in the other barcoded rabies studies. Saunders et al. found that in cell culture conditions, more than 97% of cells were infected with a single barcode and that increasing the initial multiplicity of infection 100-fold only increased the observed average number of barcodes per starter cell from 1.07 to 1.6

(Saunders et al., 2021). Zhang et al. likewise found that 95.1% of cells contained a single barcode sequence (Zhang et al., 2024) and Lin et al. found that 76% of multiply barcoded starter cells only transferred one barcode to presynaptic cells (Lin et al., 2024). This may also have the beneficial effect of reducing the amount of successful transsynaptic infections that occur between starter cells. As some time is required for rabies to establish in each starter cell before spreading to presynaptic cells, the superinfection exclusion effect could block infection of an already infected second starter cell. However, this would not prevent barcoded rabies from spreading from an infected starter cell to an initially uninfected starter. Thus, this suggests that an experimental regime where starter cells are sufficiently sparsified but saturated by initial rabies infections from the injection would reduce the probability of transsynaptic spread between starters.

Non-neuronal infection

Whilst rabies viruses are neurotropic (Ugolini, 1995), some infection of non-neuronal cells with rabies virus has been observed (Patiño et al., 2022, 2024; Potratz et al., 2020; Ray et al., 1997), and this was also observed with barcoded rabies. Saunders et al. observed from scRNAseq that 15% of the cultured cells infected with rabies were astrocytes, and 5% were polydendrocytes (Saunders et al., 2021). Lin et al. also observed with scRNAseq that 13.1% of infected cells dissociated from *in vivo* infection of hippocampus CA1 were non-neuronal (Lin et al., 2024). From my own data I observed a lower rate of non-neuronal rabies infection, with 3.8% of cells being identified by in situ sequencing as astrocytes. This could be due to my use of the CVS-N2c strain of rabies virus, which is known to have a stronger neurotropism than SAD-B19 derivatives that were used by Lin et al. and Saunders et al. (Reardon et al., 2016). It is possible that these rabies-infected non-neuronal cells that I detected are the result of the aforementioned cell segmentation issues and poor-quality cell clustering. The lack of astrocyte gene markers in my neuron-specific marker gene panel further reduces the accuracy to which cells can be assigned to this cell type.

Barcoded SAD-B19 strain rabies viruses were also used to directly study microglia-astrocyte interactions (Clark et al., 2021). This unconventional use of rabies virus for studying non-neuronal interactions contradicts the conventionally synapse-specific nature of rabies viral spread. It is, however, not infeasible given the observations of these other barcoded rabies studies, and the fact that some strains of rabies viruses can produce abortive or even persistent infections of astrocytes *in vivo* (Tian et al., 2018). Furthermore, Clark et al. mediated astrocyte infections by direct EnvA/TVA starter infection, so spread from neurons to astrocytes was not required. Ultimately, this use of rabies to study non-neuronal interactions, presumably due to close appositions between these non-neuronal cells, cannot be directly compared to the neuron-specific connectomic considerations that have been addressed in my thesis.

Infected cell recovery rates

Other approaches using barcoded rabies have also used a scRNAseq readout of barcode sequences to achieve single-cell resolution (Saunders et al., 2021; Shin et al., 2024). Saunders et al. describe methods for producing and analysing diverse barcode libraries to maximise the percentage of starter infections that are uniquely barcoded. Whilst the scRNAseq approach to detecting rabies barcodes simplifies the readout process from these experiments, the use of dissociation and FACS sorting to collect and detect rabies-positive neurons meant that an estimated 90% of all infected cells were lost before barcode detection, with starter cells being found for only 3.7% of the remaining barcodes. These experiments were carried out in cell culture where the recovery rate of dissociated neurons should be higher than that of whole brain tissue. Shin et al. aimed to mitigate whole-cell dissociation losses by designing nuclear-targeted rabies barcode cassettes to be used with single-nuclear RNA sequencing and FANS recovery (Shin et al., 2024), but no estimates were made of the percentage of retained nuclei after dissociation. This means that a direct comparison to the scRNAseq cell retention rate cannot be drawn to make conclusions on the number of undetected shared-barcode starter cells from these experiments.

There are several remaining considerations that must be addressed with regard to the proof-of-principle experiments conducted during this thesis. Firstly, the detection of barcode sequences shared in many non-starter cells without the detection of a corresponding barcoded starter cell (Figure 5.13), demonstrates that not all starter cells which initiated presynaptic infection were detected under these experimental conditions. As discussed in Chapter 5.5, this could be due to several factors, including limitations of barcode detection sensitivity, potential limitations of glycoprotein-positive cell detection sensitivity, failure of cell segmentation resulting in unattributed starter cells, or rabies cytotoxicity related starter cell death. Determining the relative contributions of these factors and making experimental improvements to detection efficiency, cell segmentation and cell survival rate would decrease the percentage of connectivity that is unrecoverable from current barcoded connectomic experiments.

Improvements to the detection of starter cells would also aid in minimising the ambiguity of connectivity resulting from barcodes shared between multiple starter cells. In the dataset described here, 27% of the barcodes detected in starter cells were present in more than one starter cell (Figure 5.12). Whilst starter cell density was titrated with the use of sparse AAV-PHP.eB-Cre infection paired with Cre-dependent helper AAVs, the density observed in this mouse was above the targeted density. This was likely due to the variability of AAV tail vein injections. The simplest way to reduce the rate of transsynaptic spread in future experiments will be to perform sequencing with a lower density of starter cells. The starter cell density can be estimated before committing to sequencing, as the mCherry fluorescence marker of glycoprotein-positive cells is imaged before sequencing begins. The variability of Cre sparsity could also be reduced in future

experiments through the use of retro-orbital injections (Prabhakar et al., 2021; Tang et al., 2024). Increasing the density of starter infections without increasing the probability of transsynaptic spread between starter cells could enable denser mapping of neural circuits with rabies viruses. This would require a mechanism for preventing transsynaptic rabies infection of starter cells after the initial rabies injection. One possibility for achieving this would be to use the AAV helper viruses that define starter cells with glycoprotein expression to also selectively block non-EnvA mediated rabies infections in starter cells. This may be achievable by knocking out the cell surface receptors that rabies interacts with during infection. It is proposed that rabies infection is mediated by multiple paths, including rabies glycoprotein interactions with the neural cell adhesion molecule NCAM (Thoulouze et al., 1998), the p75 neurotrophin receptor (p75NTR) (Tuffereau et al., 1998) and the mGluR2 receptor (Wang, Wang, et al., 2018). Reducing the expression of one or more of these receptors in starter cells may reduce the rate of transsynaptic spread sufficiently for denser barcoded rabies connectomics experiments.

6.2 Conclusion and Future Directions

The work I have described in this thesis represents a step towards a scalable, cost-effective method that can bridge the gap between transcriptomic, connectomic, spatial and functional analyses of neural circuits at the single-cell level. I have developed the required components for performing connectomic experiments, and with the preliminary data described here, I have highlighted the capability of barcoded rabies viruses and the necessary improvements to this connectomic method that will enable future experiments to determine connectivity in a robust manner.

Many avenues of research could benefit from the ability to incorporate combined connectomic and transcriptomic information. With the throughput that this method enables, comparative studies of connectivity become possible. This could allow for the study of connectomic changes in disease. The progressive loss of neuronal connectivity following the onset of dementia has led to the concept of the ”disconnectome” (Liang et al., 2023; Mandino et al., 2024), which describes a pattern of progressive disconnectivity. This loss of connectivity has already been studied in a bulk manner using monosynaptic rabies tracing in mouse models of Alzheimer’s disease (Ye et al., 2022). Single-cell connectivity changes could be studied across cell types at different time points or with the application of therapies for a variety of disease models. The compatibility of rabies viruses with infecting mammalian brains could also allow for comparative connectomic studies between species to study conserved and diverging connectivity rules throughout evolution.

Besides these applications of comparative connectomics, the use of in situ sequencing to detect barcoded rabies and endogenous gene transcripts also provides an opportunity to incorporate further modalities into a single experiment. The combination of spatial

transcriptomics and functional imaging of neuronal populations has recently been demonstrated as a powerful tool for furthering our understanding of the roles that particular cell types play in neural circuits (Bugeon et al., 2022). The ability to study the connectivity of functionally defined cells beyond the limited volume range of electron microscopy-based functional-connectomic studies (Bock et al., 2011; Lee et al., 2016; MICrONS Consortium et al., 2021) would already open up many new lines of questioning involving long-range connectivity. Combining this with transcriptomic cell typing would provide another level of information that is not accessible from EM studies. Finally, the most conceptually simple application of this high-throughput method for studying connectivity is to apply it to the plethora of regions in the central nervous system and beyond where rabies can provide new insights into the structure of diverse neural circuitries.

Chapter 7

Appendix

Table 7.1: Beamsplitter parts list

Thorlabs part name	Description	Quantity
SM2NFM	Female F-mount to SM2 external	1
SM2M35	SM2 extension tube	1
SM2V15	Adjustable SM2 internal	1
SM2L20	Stackable fixed SM2	2
SM2A20	SM2 to M38 adapter	5
TTL200	200mm tube lens	5
SM1A71	External SM2 to external SM1 adapter	1
C4W	Cube	3
C4W-CC	Cube connector	2
B4CRP/M	Rotating mirror mount	3
FFM1	Mirror spring clamp	3
B1C/M	Blank cube cover plate	3
SM1CP2	SM1 threaded caps	3
ER05-P4	Cage rod	8
CP33T/M	0.5” thick 30mm cage plate	1
CFH2R/M	Filter holder	4
SM1A71	External SM2 to external SM1 adapter	5
SM2M30	Stackable SM2 tube	4
SM2V10	Adjustable SM2	4
LCP8S	Cage plate	4
LCP16R2/M	Rotation mount	4
SM2T2	SM2 external medium length	8
TMA4	M42 external to SM2 internal	4
RSH2/M	25mm post housing	4
RS50P4/M	25mm post - M4 - 50mm long	4
PF85B-P5	Post clamping fork	1

Table 7.2: Library sequencing primer table

Primer	**Sequence**
UMI_P5_Trueseq	CTACACGACGCTCTTCC-GATCTNNNNNNNNNNNNGAAATGCCCTGAGTCCACCC
P7_universal_reverse_primer	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACGAC-CATGTCTGATCAACTG
IDT_i5_X	AATGATACGGCGACCACCGAGATCTACAC-NNNNNNNNACACTCTTTCCCTACACGACGCTCTTCC-GATC
IDT_i7_X	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGT-GACTGGAGTTCAGACGTGTG
Barcode_F	GATCTGCGGCC-GACTGC(N:25252525)(N)(N)(N)(N)(N)(N)(N)(N)GGCGCC-TAGCCGGTCATCC
Barcode_R	ATGCAGCGGCCGAC-CGG(N:25252525)(N)(N)(N)(N)(N)(N)(N)(N)CCGGGGTG-GACTCAGGGC
Supercutter_F2	AGCCAGGCGGGGATCCGGCTGC
Supercutter_R	CTCCCTTAGCCATCCGAGTGGACG

Table 7.4: In situ sequencing primer table

Primer	**Sequence**
PAB0010	GGCATCTAACGGCTCTTAGTCAACTCTAC
PAB0012	GCAATAAGTGCCGCGTGCTCGACG
PAB0013	GACGATCTGTCGACAACTATTAGTACAGCG
PAB0022	CGACAGGTTCAGAGTTCTACAGTCCGACGAT
PAB0025	/5ATTO565N/AGTCAGCGTCGAGCACGCGGCACT-TATTGCA
PAB0026	/5ATTO532N/ATTGCACGATCCAGTTGCTGACCGCTTCAG
PAB0027	/5ATTO633N/TCGCTGTACTAATAGTTGTCGACA-GATCGTCA
PAB0028	/5ATTO700N/TGAGTAGAGTTGACTAAGAGCCGTTAGAT-GCC
PAB0029	/5ATTO565N/CCTCTATGATTACTGACTGCGTCTA
PAB0030	/5ATTO565N/ATCGTCGGACTGTAGAACTCTGAACCT-GTCG
PAB0032	A+TCG+TCG+GAC+TGT+AGA+ACT+CTGAACCTGTCG
PAB0033	GAAGCGGTCAGCAACTGGATCGTG
PAB0034	TTGCTATCTTCTTTCCACCCCGGCCGACTGC
PAB0035	GGCCGGGGTGGAAAGAAGATAGCAATG
OAB0973	/5Phos/GGCGCCTAGCCGGTCATCCTTCCTCTATGAT-TACTGACTGCGTCTATTTAGTGGAGC ATTGCTATCATTGCTATCTTCTTTCCACCCCGGCCGACTGC

Bibliography

Adachi, K., Enoki, T., Kawano, Y., Veraz, M., & Nakai, H. (2014). Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing [Publisher: Nature Publishing Group]. *Nature Communications*, *5*(1), 3075. https://doi.org/10.1038/ncomms4075

Albers, J., Svetlove, A., & Duke, E. (2024). Synchrotron X-ray imaging of soft biological tissues - principles, applications and future prospects. *Journal of Cell Science*, *137*(20), jcs261953. https://doi.org/10.1242/jcs.261953

Albertini, A. A. V., Ruigrok, R. W. H., & Blondel, D. (2011). Rabies virus transcription and replication. *Advances in Virus Research*, *79*, 1–22. https://doi.org/10.1016/B978-0-12-387040-7.00001-9

Alon, S., Goodwin, D. R., Sinha, A., Wassie, A. T., Chen, F., Daugharthy, E. R., Bando, Y., Kajita, A., Xue, A. G., Marrett, K., Prior, R., Cui, Y., Payne, A. C., Yao, C.-C., Suk, H.-J., Wang, R., Yu, C.-C. (, Tillberg, P., Reginato, P., … Zhuang, X. (2021). Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science*, *371*(6528), eaax2656. https://doi.org/10.1126/science.aax2656

Archin, N. M., & Atherton, S. S. (2002). Rapid spread of a neurovirulent strain of HSV-1 through the CNS of BALB/c mice following anterior chamber inoculation. *Journal of Neurovirology*, *8*(2), 122–135. https://doi.org/10.1080/13550280290049570

Astic, L., Saucier, D., & Flamand, A. (1993). The CVS strain of rabies virus as transneuronal tracer in the olfactory system of mice [Publisher: Elsevier]. *Brain Research*, *619*(1-2), 146–156. https://doi.org/10.1016/0006-8993(93)91606-S

Aston-Jones, G., & Card, J. P. (2000). Use of pseudorabies virus to delineate multisynaptic circuits in brain: Opportunities and limitations. *Journal of Neuroscience Methods*, *103*(1), 51–61. https://doi.org/10.1016/s0165-0270(00)00295-8

Battich, N., Stoeger, T., & Pelkmans, L. (2013). Image-based transcriptomics in thousands of single human cells at single-molecule resolution [Publisher: Nature Publishing Group]. *Nature Methods*, *10*(11), 1127–1133. https://doi.org/10.1038/nmeth.2657

Bauer, A., Nolden, T., Schröter, J., Römer-Oberdörfer, A., Gluska, S., Perlson, E., & Finke, S. (2014). Anterograde Glycoprotein-Dependent Transport of Newly Generated Rabies Virus in Dorsal Root Ganglion Neurons [Publisher: American Society for Microbiology]. *Journal of Virology*, *88*(24), 14172–14183. https://doi.org/10.1128/jvi.02254-14

Bhang, H.-e. C., Ruddy, D. A., Krishnamurthy Radhakrishna, V., Caushi, J. X., Zhao, R., Hims, M. M., Singh, A. P., Kao, I., Rakiec, D., Shaw, P., Balak, M., Raza, A., Ackley, E., Keen, N., Schlabach, M. R., Palmer, M., Leary, R. J., Chiang, D. Y., Sellers, W. R., … Stegmeier, F. (2015). Studying clonal dynamics in response to cancer therapy using high-complexity barcoding [Publisher: Nature Publishing Group]. *Nature Medicine*, *21*(5), 440–448. https://doi.org/10.1038/nm.3841

Bock, D. D., Lee, W.-C. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., Yurgenson, S., Soucy, E. R., Kim, H. S., & Reid, R. C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, *471*(7337), 177–182. https://doi.org/10.1038/nature09802

Bogovic, J. A., Hanslovsky, P., Wong, A., & Saalfeld, S. (2016). Robust registration of calcium images by learned contrast synthesis [ISSN: 1945-8452]. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 1123–1126. https://doi.org/10.1109/ISBI.2016.7493463

Bosch, C., Ackels, T., Pacureanu, A., Zhang, Y., Peddie, C. J., Berning, M., Rzepka, N., Zdora, M.-C., Whiteley, I., Storm, M., Bonnin, A., Rau, C., Margrie, T., Collinson, L., & Schaefer, A. T. (2022). Functional and multiscale 3D structural investigation of brain tissue through correlative in vivo physiology, synchrotron microtomography and volume electron microscopy. *Nature Communications*, *13*, 2923. https://doi.org/10.1038/s41467-022-30199-6

Bosch, C., Diaz, A., Holler, M., Guizar-Sicairos, M., Aidukas, T., Pacureanu, A., Müller, E., Peddie, C. J., Collinson, L., Zhang, Y., Menzel, A., Wanner, A. A., & Schaefer, A. T. (2023, November). 3D-Imaging of synapses in neuronal tissues with synchrotron X-ray ptychography [Pages: 2023.11.16.567403 Section: New Results]. https://doi.org/10.1101/2023.11.16.567403

Brown, D., Altermatt, M., Dobreva, T., Chen, S., Wang, A., Thomson, M., & Gradinaru, V. (2021). Deep Parallel Characterization of AAV Tropism and AAV-Mediated Transcriptional Changes via Single-Cell RNA Sequencing [Publisher: Frontiers]. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.730825

Bugeon, S., Duffield, J., Dipoppa, M., Ritoux, A., Prankerd, I., Nicoloutsopoulos, D., Orme, D., Shinn, M., Peng, H., Forrest, H., Viduolyte, A., Reddy, C. B., Isogai, Y., Carandini, M., & Harris, K. D. (2022). A transcriptomic axis predicts state modulation of cortical interneurons [Publisher: Nature Publishing Group]. *Nature*, *607*(7918), 330–338. https://doi.org/10.1038/s41586-022-04915-7

Bumbarger, D. J., Riebesell, M., Rödelsperger, C., & Sommer, R. J. (2013). System-wide rewiring underlies behavioral differences in predatory and bacterial-feeding nematodes. *Cell*, *152*(1-2), 109–119. https://doi.org/10.1016/j.cell.2012.12.013

Büttner-Ennever, J. A., Grob, P., Akert, K., & Bizzini, B. (1981). A transsynaptic autoradiographic study of the pathways controlling the extraocular eye muscles, using [125I]B-IIb tetanus toxin fragment [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1981.tb30868.x]. *Annals of the New York Academy of Sciences*, *374*(1), 157–170. https://doi.org/10.1111/j.1749-6632.1981.tb30868.x

Cadwell, C. R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K. F., Sandberg, R., & Tolias, A. S. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nature Biotechnology*, *34*(2), 199–203. https://doi.org/10.1038/nbt.3445

Callaway, E. M., & Luo, L. (2015). Monosynaptic Circuit Tracing with Glycoprotein-Deleted Rabies Viruses [Publisher: Society for Neuroscience Section: Toolbox]. *Journal of Neuroscience*, *35*(24), 8979–8985. https://doi.org/10.1523/JNEUROSCI.0409-15.2015

Campagnola, L., Seeman, S. C., Chartrand, T., Kim, L., Hoggarth, A., Gamlin, C., Ito, S., Trinh, J., Davoudian, P., Radaelli, C., Kim, M.-H., Hage, T., Braun, T., Alfiler, L., Andrade, J., Bohn, P., Dalley, R., Henry, A., Kebede, S., … Jarsky, T. (2022). Local connectivity and synaptic dynamics in mouse and human neocortex. *Science*, *375*(6585), eabj5861. https://doi.org/10.1126/science.abj5861

Campbell, R. (2020, January). SainsburyWellcomeCentre/BakingTray: Jan 2020. https://doi.org/10.5281/zenodo.3631610

Campbell, R., Blot, A., & lguerard. (2020, July). SainsburyWellcomeCentre/StitchIt: Last release of stitching model 1. https://doi.org/10.5281/zenodo.3941901

Card, J. P., Levitt, P., & Enquist, L. W. (1998). Different Patterns of Neuronal Infection after Intracerebral Injection of Two Strains of Pseudorabies Virus. *Journal of Virology*, *72*(5), 4434–4441.

Carey, H., Pegios, M., Martin, L., Saleeba, C., Turner, A. J., Everett, N. A., Bjerke, I. E., Puchades, M. A., Bjaalie, J. G., & McMullan, S. (2023). DeepSlice: Rapid fully automatic registration

of mouse brain imaging to a volumetric atlas [Publisher: Nature Publishing Group]. *Nature Communications*, *14*(1), 5884. https://doi.org/10.1038/s41467-023-41645-4

Chang, J.-B., Chen, F., Yoon, Y.-G., Jung, E. E., Babcock, H., Kang, J. S., Asano, S., Suk, H.-J., Pak, N., Tillberg, P. W., Wassie, A. T., Cai, D., & Boyden, E. S. (2017). Iterative expansion microscopy. *Nature Methods*, *14*(6), 593–599. https://doi.org/10.1038/nmeth.4261

Chang, T., Han, W., Jiang, M., Li, J., Liao, Z., Tang, M., Zhang, J., Shen, J., Chen, Z., Fei, P., Ren, X., Pang, Y., Wang, G., Wang, J., & Huang, Y. (2023). Rapid and signal crowdedness-robust in situ sequencing through hybrid block coding [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *120*(47), e2309227120. https://doi.org/10.1073/pnas.2309227120

Chatterjee, D., Marmion, D. J., McBride, J. L., Manfredsson, F. P., Butler, D., Messer, A., & Kordower, J. H. (2022). Enhanced CNS transduction from AAV.PHP.eB infusion into the cisterna magna of older adult rats compared to AAV9 [Publisher: Nature Publishing Group]. *Gene Therapy*, *29*(6), 390–397. https://doi.org/10.1038/s41434-021-00244-y

Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., Wang, X., Lu, H., Chen, X., Liu, X., Huang, X., Li, Z., Hong, Y., Jiang, Y., Peng, J., … Wang, J. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays [Publisher: Elsevier]. *Cell*, *185*(10), 1777–1792.e21. https://doi.org/10.1016/j.cell.2022.04.003

Chen, F., Wassie, A. T., Cote, A. J., Sinha, A., Alon, S., Asano, S., Daugharthy, E. R., Chang, J.-B., Marblestone, A., Church, G. M., Raj, A., & Boyden, E. S. (2016). Nanoscale imaging of RNA with expansion microscopy. *Nature Methods*, *13*(8), 679–684. https://doi.org/10.1038/nmeth.3899

Chen, H., Li, D., & Bar-Joseph, Z. (2023). SCS: Cell segmentation for high-resolution spatial transcriptomics [Publisher: Nature Publishing Group]. *Nature Methods*, *20*(8), 1237–1243. https://doi.org/10.1038/s41592-023-01939-3

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells [Publisher: American Association for the Advancement of Science]. *Science*, *348*(6233), aaa6090. https://doi.org/10.1126/science.aaa6090

Chen, R., Wu, X., Jiang, L., & Zhang, Y. (2017). Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity [Publisher: Elsevier]. *Cell Reports*, *18*(13), 3227–3241. https://doi.org/10.1016/j.celrep.2017.03.004

Chen, S., Loper, J., Chen, X., Vaughan, A., Zador, A. M., & Paninski, L. (2021). BARcode DEmixing through Non-negative Spatial Regression (BarDensr) (E. C. Robinson, Ed.). *PLOS Computational Biology*, *17*(3), e1008256. https://doi.org/10.1371/journal.pcbi.1008256

Chen, X., Fischer, S., Rue, M. C. P., Zhang, A., Mukherjee, D., Kanold, P. O., Gillis, J., & Zador, A. M. (2024). Whole-cortex in situ sequencing reveals input-dependent area identity [Publisher: Nature Publishing Group]. *Nature*, 1–10. https://doi.org/10.1038/s41586-024-07221-6

Chen, X., Sun, Y.-C., Church, G. M., Lee, J. H., & Zador, A. M. (2018). Efficient in situ barcode sequencing using padlock probe-based BaristaSeq [BaristaSeq]. *Nucleic Acids Research*, *46*(4), e22–e22. https://doi.org/10.1093/nar/gkx1206

Chen, X., Sun, Y.-C., Zhan, H., Kebschull, J. M., Fischer, S., Matho, K., Huang, Z. J., Gillis, J., & Zador, A. M. (2019). High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing [BARseq]. *Cell*, *179*(3), 772–786.e19. https://doi.org/10.1016/j.cell.2019.09.023

Chen, Y., Xu, X., Wan, X., Xiao, J., & Yang, C. (2024, July). UCS: A unified approach to cell segmentation for subcellular spatial transcriptomics [Pages: 2024.07.08.601384 Section: New Results]. https://doi.org/10.1101/2024.07.08.601384

Chen, Y., Chen, X., Baserdem, B., Zhan, H., Li, Y., Davis, M. B., Kebschull, J. M., Zador, A. M., Koulakov, A. A., & Albeanu, D. F. (2022). High-throughput sequencing of single neuron projections reveals spatial organization in the olfactory cortex [Publisher: Elsevier]. *Cell*, *185*(22), 4117–4134.e28. https://doi.org/10.1016/j.cell.2022.09.038

Cherng, J. .-., Schuurmans-Nieuwenbroek, N. M. E., Jiskoot, W., Talsma, H., Zuidam, N. J., Hennink, W. E., & Crommelin, D. J. A. (1999). Effect of DNA topology on the transfection efficiency of poly((2-dimethylamino)ethyl methacrylate)–plasmid complexes. *Journal of Controlled Release*, *60*(2), 343–353. https://doi.org/10.1016/S0168-3659(99)00089-9

Cheung, V., Chung, P., Bjorni, M., Shvareva, V. A., Lopez, Y. C., & Feinberg, E. H. (2021). Virally encoded connectivity transgenic overlay RNA sequencing (VECTORseq) defines projection neurons involved in sensorimotor integration [Publisher: Elsevier]. *Cell Reports*, *37*(12). https://doi.org/10.1016/j.celrep.2021.110131

Chiaruttini, N., Castoldi, C., Requie, L. M., Camarena-Delgado, C., dal Bianco, B., Gräff, J., Seitz, A., & Silva, B. (2024). ABBA, a novel tool for whole-brain mapping, reveals brain-wide differences in immediate early genes induction following learnin. https://doi.org/10.1101/2024.09.06.611625

Cho, J.-H., Deisseroth, K., & Bolshakov, V. Y. (2013). Synaptic Encoding of Fear Extinction in mPFC-amygdala Circuits [Publisher: Elsevier]. *Neuron*, *80*(6), 1491–1507. https://doi.org/10.1016/j.neuron.2013.09.025

Choi, H. M. T., Beck, V. A., & Pierce, N. A. (2014). Next-Generation in Situ Hybridization Chain Reaction: Higher Gain, Lower Cost, Greater Durability [Publisher: American Chemical Society]. *ACS Nano*, *8*(5), 4284–4294. https://doi.org/10.1021/nn405717p

Clark, I. C., Gutiérrez-Vázquez, C., Wheeler, M. A., Li, Z., Rothhammer, V., Linnerbauer, M., Sanmarco, L. M., Guo, L., Blain, M., Zandee, S. E. J., Chao, C.-C., Batterman, K. V., Schwabenland, M., Lotfy, P., Tejeda-Velarde, A., Hewson, P., Polonio, C. M., Shultis, M. W., Salem, Y., … Quintana, F. J. (2021). Barcoded viral tracing of single-cell interactions in central nervous system inflammation [Publisher: American Association for the Advancement of Science Section: Research Article]. *Science*, *372*(6540). https://doi.org/10.1126/science.abf1230

Claudi, F., Petrucco, L., Tyson, A. L., Branco, T., Margrie, T. W., & Portugues, R. (2020). BrainGlobe Atlas API: A common interface for neuroanatomical atlases. *Journal of Open Source Software*, *5*(54), 2668. https://doi.org/10.21105/joss.02668

Combe, M., Garijo, R., Geller, R., Cuevas, J. M., & Sanjuán, R. (2015). Single-Cell Analysis of RNA Virus Infection Identifies Multiple Genetically Diverse Viral Genomes within Single Infectious Units [Publisher: Elsevier]. *Cell Host & Microbe*, *18*(4), 424–432. https://doi.org/10.1016/j.chom.2015.09.009

Conzelmann, K. K. (2004). Reverse Genetics of Mononegavirales. In Y. Kawaoka (Ed.), *Biology of Negative Strand RNA Viruses: The Power of Reverse Genetics* (pp. 1–41). Springer. https://doi.org/10.1007/978-3-662-06099-5_1

Conzelmann, K. K., & Schnell, M. (1994). Rescue of synthetic genomic RNA analogs of rabies virus by plasmid-encoded proteins. *Journal of Virology*, *68*(2), 713–719. https://doi.org/10.1128/JVI.68.2.713-719.1994

Conzelmann, K.-K. 1. (1996). Genetic manipulation of non-segmented negative-strand RNA viruses [Publisher: Microbiology Society,]. *The Journal of General Virology*, *77 (Pt 3)*(3), 381–389. https://doi.org/10.1099/0022-1317-77-3-381

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., & Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex [Number: 7539 Publisher: Nature Publishing Group]. *Nature*, *518*(7539), 399–403. https://doi.org/10.1038/nature14182

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., Hayden Gephart, M. G., Barres, B. A., & Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *112*(23), 7285–7290. https://doi.org/10.1073/pnas.1507125112

Davidsson, M., Diaz-Fernandez, P., Schwich, O. D., Torroba, M., Wang, G., & Björklund, T. (2016). A novel process of viral vector barcoding and library preparation enables high-diversity library

generation and recombination-free paired-end sequencing [Publisher: Nature Publishing Group]. *Scientific Reports*, *6*(1), 37563. https://doi.org/10.1038/srep37563

Deng, R., Zhang, K., Wang, L., Ren, X., Sun, Y., & Li, J. (2018). DNA-Sequence-Encoded Rolling Circle Amplicon for Single-Cell RNA Imaging. *Chem*, *4*(6), 1373–1386. https://doi.org/10.1016/j.chempr.2018.03.003

Denk, W., & Horstmann, H. (2004). Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLoS Biology*, *2*(11), e329. https://doi.org/10.1371/journal.pbio.0020329

Dirks, R. M., & Pierce, N. A. (2004). Triggered amplification by hybridization chain reaction [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *101*(43), 15275–15278. https://doi.org/10.1073/pnas.0407024101

Donnelly, M. L. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., & Ryan, M. D. (2001). The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences [Publisher: Microbiology Society,]. *Journal of General Virology*, *82*(5), 1027–1041. https://doi.org/10.1099/0022-1317-82-5-1027

Dorkenwald, S., Turner, N. L., Macrina, T., Lee, K., Lu, R., Wu, J., Bodor, A. L., Bleckert, A. A., Brittain, D., Kemnitz, N., Silversmith, W. M., Ih, D., Zung, J., Zlateski, A., Tartavull, I., Yu, S.-C., Popovych, S., Wong, W., Castro, M., … Seung, H. S. (2022). Binary and analog variation of synapses between cortical pyramidal neurons (T. O'Leary, J. R. Huguenard, H. Kasai, & T. M. Bartol, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *11*, e76120. https://doi.org/10.7554/eLife.76120

Dorkenwald, S., Matsliah, A., Sterling, A. R., Schlegel, P., Yu, S.-c., McKellar, C. E., Lin, A., Costa, M., Eichler, K., Yin, Y., Silversmith, W., Schneider-Mizell, C., Jordan, C. S., Brittain, D., Halageri, A., Kuehner, K., Ogedengbe, O., Morey, R., Gager, J., … Murthy, M. (2024). Neuronal wiring diagram of an adult brain [Publisher: Nature Publishing Group]. *Nature*, *634*(8032), 124–138. https://doi.org/10.1038/s41586-024-07558-y

Dräger, U. C. (1974). Autoradiography of tritiated proline and fucose transported transneuronally from the eye to the visual cortex in pigmented and albino mice. *Brain Research*.

Du, M., & Jacobsen, C. (2018). Relative merits and limiting factors for x-ray and electron microscopy of thick, hydrated organic materials. *Ultramicroscopy*, *184*(Pt A), 293–309. https://doi.org/10.1016/j.ultramic.2017.10.003

Economo, M. N., Clack, N. G., Lavis, L. D., Gerfen, C. R., Svoboda, K., Myers, E. W., & Chandrashekar, J. (2016). A platform for brain-wide imaging and reconstruction of individual neurons. *eLife*, *5*, e10566. https://doi.org/10.7554/eLife.10566

Economo, M. N., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L. T., Nguyen, T. N., Smith, K. A., Yao, Z., Wang, L., Gerfen, C. R., Chandrashekar, J., Zeng, H., Looger, L. L., & Svoboda, K. (2018). Distinct descending motor cortex pathways and their roles in movement [Publisher: Nature Publishing Group]. *Nature*, *563*(7729), 79–84. https://doi.org/10.1038/s41586-018-0642-9

Edelstein, A., Amodaj, N., Hoover, K., Vale, R., & Stuurman, N. (2010). Computer Control of Microscopes Using μManager [ISBN: 9780471142720 Publisher: John Wiley & Sons, Ltd]. *Current Protocols in Molecular Biology*, *92*(1), 14.20.1–14.20.17. https://doi.org/10.1002/0471142727.mb1420s92

Elabbady, L., Seshamani, S., Mu, S., Mahalingam, G., Schneider-Mizell, C., Bodor, A. L., Bae, J. A., Brittain, D., Buchanan, J., Bumbarger, D. J., Castro, M. A., Dorkenwald, S., Halageri, A., Jia, Z., Jordan, C., Kapner, D., Kemnitz, N., Kinn, S., Lee, K., … Collman, F. (2024, January). Perisomatic Features Enable Efficient and Dataset Wide Cell-Type Classifications Across Large-Scale Electron Microscopy Volumes [Pages: 2022.07.20.499976 Section: New Results]. https://doi.org/10.1101/2022.07.20.499976

Eng, C.-H. L., Lawson, M., Zhu, Q., Dries, R., Koulena, 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+ [Publisher: Nature Publishing Group]. *Nature*, *568*(7751), 235–239. https://doi.org/10.1038/s41586-019-1049-y

Etessami, R., Conzelmann, K.-K., Fadai-Ghotbi, B., Natelson, B., Tsiang, H., & Ceccaldi, P.-E. (2000). Spread and pathogenic characteristics of a G-deficient rabies virus recombinant: An in vitro and in vivo study. *The Journal of General Virology*, *81*(Pt 9), 2147–2153. https://doi.org/10.1099/0022-1317-81-9-2147

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 [Publisher: American Association for the Advancement of Science]. *Science*, *377*(6601), 56–62. https://doi.org/10.1126/science.abm1741

Fennell, K. A., Vassiliadis, D., Lam, E. Y. N., Martelotto, L. G., Balic, J. J., Hollizeck, S., Weber, T. S., Semple, T., Wang, Q., Miles, D. C., MacPherson, L., Chan, Y.-C., Guirguis, A. A., Kats, L. M., Wong, E. S., Dawson, S.-J., Naik, S. H., & Dawson, M. A. (2022). Non-genetic determinants of malignant clonal fitness at single-cell resolution [Publisher: Nature Publishing Group]. *Nature*, *601*(7891), 125–131. https://doi.org/10.1038/s41586-021-04206-7

Finke, S., Cox, J. H., & Conzelmann, K.-K. (2000). Differential Transcription Attenuation of Rabies Virus Genes by Intergenic Regions: Generation of Recombinant Viruses Overexpressing the Polymerase Gene [Publisher: American Society for Microbiology]. *Journal of Virology*, *74*(16), 7261–7269. https://doi.org/10.1128/jvi.74.16.7261-7269.2000

Fischer, K. B., Collins, H. K., & Callaway, E. M. (2019). Sources of off-target expression from recombinase-dependent AAV vectors and mitigation with cross-over insensitive ATG-out vectors. *Proceedings of the National Academy of Sciences*, *116*(52), 27001–27010. https://doi.org/10.1073/pnas.1915974116

Fuzik, J., Zeisel, A., Máté, Z., Calvigioni, D., Yanagawa, Y., Szabó, G., Linnarsson, S., & Harkany, T. (2016). Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nature Biotechnology*, *34*(2), 175–183. https://doi.org/10.1038/nbt.3443

Galarreta, M., & Hestrin, S. (2002). Electrical and chemical synapses among parvalbumin fast-spiking GABAergic interneurons in adult mouse neocortex [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *99*(19), 12438–12443. https://doi.org/10.1073/pnas.192159599

Galloni, A. R., Ye, Z., & Rancz, E. (2022). Dendritic Domain-Specific Sampling of Long-Range Axons Shapes Feedforward and Feedback Connectivity of L5 Neurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(16), 3394–3405. https://doi.org/10.1523/JNEUROSCI.1620-21.2022

Gao, L., Liu, S., Gou, L., Hu, Y., Liu, Y., Deng, L., Ma, D., Wang, H., Yang, Q., Chen, Z., Liu, D., Qiu, S., Wang, X., Wang, D., Wang, X., Ren, B., Liu, Q., Chen, T., Shi, X., … Yan, J. (2022). Single-neuron projectome of mouse prefrontal cortex. *Nature Neuroscience*, *25*(4), 515–529. https://doi.org/10.1038/s41593-022-01041-5

Gao, R., Asano, S. M., Upadhyayula, S., Pisarev, I., Milkie, D. E., Liu, T.-L., Singh, V., Graves, A., Huynh, G. H., Zhao, Y., Bogovic, J., Colonell, J., Ott, C. M., Zugates, C., Tappan, S., Rodriguez, A., Mosaliganti, K. R., Sheu, S.-H., Pasolli, H. A., … Betzig, E. (2019). Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. *Science (New York, N.Y.)*, *363*(6424). https://doi.org/10.1126/science.aau8302

Geddes, S. D., Assadzada, S., Lemelin, D., Sokolovski, A., Bergeron, R., Haj-Dahmane, S., & Béïque, J.-C. (2016). Target-specific modulation of the descending prefrontal cortex inputs to the dorsal raphe nucleus by cannabinoids. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(19), 5429–5434. https://doi.org/10.1073/pnas.1522754113

Gergues, M. M., Han, K. J., Choi, H. S., Brown, B., Clausing, K. J., Turner, V. S., Vainchtein, I. D., Molofsky, A. V., & Kheirbek, M. A. (2020). Circuit and molecular architecture of a ventral hippocampal network. *Nature Neuroscience*, *23*(11), 1444–1452. https://doi.org/10.1038/s41593-020-0705-8

Ghanem, A., Kern, A., & Conzelmann, K.-K. (2012). Significantly improved rescue of rabies virus from cDNA plasmids. *European Journal of Cell Biology*, *91*(1), 10–16. https://doi.org/10.1016/j.ejcb.2011.01.008

Gilbert, C. D., & Wiesel, T. N. (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature*, *280*(5718), 120–125. https://doi.org/10.1038/280120a0

Gonatas, N. K., Harper, C., Mizutani, T., & Gonatas, J. O. (1979). Superior sensitivity of conjugates of horseradish peroxidase with wheat germ agglutinin for studies of retrograde axonal transport. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, *27*(3), 728–734. https://doi.org/10.1177/27.3.90065

Goodwin, D. R., Vaughan, A., Leible, D., Alon, S., Henry, G. L., Cheng, A., Chen, X., Zhang, R., Xue, A. G., Wassie, A. T., Sinha, A., Bando, Y., Kajita, A., Marblestone, A. H., Zador, A. M., Boyden, E. S., Church, G. M., & Kohman, R. E. (2022, July). *Expansion Sequencing of RNA Barcoded Neurons in the Mammalian Brain: Progress and Implications for Molecularly Annotated Connectomics* (preprint). Neuroscience. https://doi.org/10.1101/2022.07.31.502046

Gouwens, N. W., Sorensen, S. A., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S. M., Feng, D., Anastassiou, C. A., Barkan, E., Bickley, K., Blesie, N., Braun, T., Brouner, K., Budzillo, A., Caldejon, S., Casper, T., Castelli, D., Chong, P., … Koch, C. (2019). Classification of electrophysiological and morphological types in mouse visual cortex. *Nature neuroscience*, *22*(7), 1182–1195. https://doi.org/10.1038/s41593-019-0417-0

Gouwens, N. W., Sorensen, S. A., Baftizadeh, F., Budzillo, A., Lee, B. R., Jarsky, T., Alfiler, L., Baker, K., Barkan, E., Berry, K., Bertagnolli, D., Bickley, K., Bomben, J., Braun, T., Brouner, K., Casper, T., Crichton, K., Daigle, T. L., Dalley, R., … Zeng, H. (2020). Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell*, *183*(4), 935–953.e19. https://doi.org/10.1016/j.cell.2020.09.057

Graf, W., Gerrits, N., Yatim-Dhiba, N., & Ugolini, G. (2002). Mapping the oculomotor system: The power of transneuronal labelling with rabies virus [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1460-9568.2002.01994.x]. *European Journal of Neuroscience*, *15*(9), 1557–1562. https://doi.org/10.1046/j.1460-9568.2002.01994.x

Grafstein, B. (1971). Transneuronal Transfer of Radioactivity in the Central Nervous System [Publisher: American Association for the Advancement of Science]. *Science*, *172*(3979), 177–179. https://doi.org/10.1126/science.172.3979.177

Greb, C. (2016). Infinity Optical Systems. *Optik & Photonik*, *11*(1), 34–37. https://doi.org/10.1002/opph.201600002

Habel, K. (1940). Evaluation of a Mouse Test for the Standardization of the Immunizing Power of Anti-Rabies Vaccines. *Public Health Reports (1896-1970)*, *55*(33), 1473. https://doi.org/10.2307/4583406

Hafner, G., Witte, M., Guy, J., Subhashini, N., Fenno, L. E., Ramakrishnan, C., Kim, Y. S., Deisseroth, K., Callaway, E. M., Oberhuber, M., Conzelmann, K.-K., & Staiger, J. F. (2019). Mapping Brain-Wide Afferent Inputs of Parvalbumin-Expressing GABAergic Neurons in Barrel Cortex Reveals Local and Long-Range Circuit Motifs [Publisher: Elsevier]. *Cell Reports*, *28*(13), 3450–3461.e8. https://doi.org/10.1016/j.celrep.2019.08.064

Hage, T. A., Bosma-Moody, A., Baker, C. A., Kratz, M. B., Campagnola, L., Jarsky, T., Zeng, H., & Murphy, G. J. (2022). Synaptic connectivity to L2/3 of primary visual cortex measured by two-photon optogenetic stimulation. *eLife*, *11*, e71103. https://doi.org/10.7554/eLife.71103

Han, Y., Kebschull, J. M., Campbell, R. A. A., Cowan, D., Imhof, F., Zador, A. M., & Mrsic-Flogel, T. D. (2018). The logic of single-cell projections from visual cortex. *Nature*, *556*(7699), 51–56. https://doi.org/10.1038/nature26159

Hanada, K., Suzuki, Y., & Gojobori, T. (2004). A Large Variation in the Rates of Synonymous Substitution for RNA Viruses and Its Relationship to a Diversity of Viral Infection and Transmission Modes. *Molecular Biology and Evolution*, *21*(6), 1074–1080. https://doi.org/10.1093/molbev/msh109

Harris, J. A., Mihalas, S., Hirokawa, K. E., Whitesell, J. D., Choi, H., Bernard, A., Bohn, P., Caldejon, S., Casal, L., Cho, A., Feiner, A., Feng, D., Gaudreault, N., Gerfen, C. R., Graddis, N., Groblewski, P. A., Henry, A. M., Ho, A., Howard, R., … Zeng, H. (2019). Hierarchical organization of cortical and thalamic connectivity. *Nature*, *575*(7781), 195–202. https://doi.org/10.1038/s41586-019-1716-z

Hausmann, F. S., Barrett, J. M., Martin, M. E., Zhan, H., & Shepherd, G. M. G. (2022). Axonal Barcode Analysis of Pyramidal Tract Projections from Mouse Forelimb M1 and M2. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(41), 7733–7743. https://doi.org/10.1523/JNEUROSCI.1062-22.2022

He, S., Bhatt, R., Brown, C., Brown, E. A., Buhr, D. L., Chantranuvatana, K., Danaher, P., Dunaway, D., Garrison, R. G., Geiss, G., Gregory, M. T., Hoang, M. L., Khafizov, R., Killingbeck, E. E., Kim, D., Kim, T. K., Kim, Y., Klock, A., Korukonda, M., … Beechem, J. M. (2022). High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging [Publisher: Nature Publishing Group]. *Nature Biotechnology*, *40*(12), 1794–1806. https://doi.org/10.1038/s41587-022-01483-z

Herculano-Houzel, S., Watson, C. R., & Paxinos, G. (2013). Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones [Publisher: Frontiers]. *Frontiers in Neuroanatomy*, *7*. https://doi.org/10.3389/fnana.2013.00035

Holmgren, C., Harkany, T., Svennenfors, B., & Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *The Journal of Physiology*, *551*(Pt 1), 139–153. https://doi.org/10.1113/jphysiol.2003.044784

Hooks, B. M., Lin, J. Y., Guo, C., & Svoboda, K. (2015). Dual-Channel Circuit Mapping Reveals Sensorimotor Convergence in the Primary Motor Cortex [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *35*(10), 4418–4426. https://doi.org/10.1523/JNEUROSCI.3741-14.2015

Howard, D. B., Powers, K., Wang, Y., & Harvey, B. K. (2008). Tropism and toxicity of adeno-associated viral vector serotypes 1, 2, 5, 6, 7, 8, and 9 in rat neurons and glia in vitro. *Virology*, *372*(1), 24–34. https://doi.org/10.1016/j.virol.2007.10.007

Hu, H., Ma, Y., & Agmon, A. (2011). Submillisecond Firing Synchrony between Different Subtypes of Cortical Interneurons Connected Chemically But Not Electrically. *The Journal of Neuroscience*, *31*(9), 3351–3361. https://doi.org/10.1523/JNEUROSCI.4881-10.2011

Hua, Y., Laserstein, P., & Helmstaedter, M. (2015). Large-volume en-bloc staining for electron microscopy-based connectomics [Publisher: Nature Publishing Group]. *Nature Communications*, *6*(1), 7923. https://doi.org/10.1038/ncomms8923

Huang, L., Kebschull, J. M., Fürth, D., Musall, S., Kaufman, M. T., Churchland, A. K., & Zador, A. M. (2020). BRICseq Bridges Brain-wide Interregional Connectivity to Neural Activity and Gene Expression in Single Animals [BRICseq]. *Cell*, *182*(1), 177–188.e27. https://doi.org/10.1016/j.cell.2020.05.029

Inoue, K. i., Shoji, Y., Kurane, I., Iijima, T., Sakai, T., & Morimoto, K. (2003). An improved method for recovering rabies virus from cloned cDNA. *Journal of Virological Methods*, *107*(2), 229–236. https://doi.org/10.1016/s0166-0934(02)00249-5

Itaya, S., & Hoesen, G. (1982). WGA-HRP as a transneuronal marker in the visual pathways of monkey and rat [Publisher: Elsevier]. *Brain Research*, *236*(1), 199–204. https://doi.org/10.1016/0006-8993(82)90046-4

Jakt, L. M., Moriwaki, S., & Nishikawa, S. (2013). A continuum of transcriptional identities visualized by combinatorial fluorescent in situ hybridization. *Development*, *140*(1), 216–225. https://doi.org/10.1242/dev.086975

Jang, M. J., Coughlin, G. M., Jackson, C. R., Chen, X., Chuapoco, M. R., Vendemiatti, J. L., Wang, A. Z., & Gradinaru, V. (2023). Spatial transcriptomics for profiling the tropism of viral vectors in tissues. *Nature Biotechnology*, *41*(9), 1272–1286. https://doi.org/10.1038/s41587-022-01648-w

Jiang, X., Shen, S., Cadwell, C. R., Berens, P., Sinz, F., Ecker, A. S., Patel, S., & Tolias, A. S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, *350*(6264), aac9462–aac9462. https://doi.org/10.1126/science.aac9462

Johnson, R. M., & Spear, P. G. (1989). Herpes simplex virus glycoprotein D mediates interference with herpes simplex virus infection. *Journal of Virology*, *63*(2), 819–827. https://doi.org/10.1128/JVI.63.2.819-827.1989

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening [Number: 4 Publisher: Nature Publishing Group]. *Nature Protocols*, *12*(4), 828–863. https://doi.org/10.1038/nprot.2017.016

Ju, J., Kim, D. H., Bi, L., Meng, Q., Bai, X., Li, Z., Li, X., Marma, M. S., Shi, S., Wu, J., Edwards, J. R., Romu, A., & Turro, N. J. (2006). Four-Color DNA Sequencing by Synthesis Using Cleavable Fluorescent Nucleotide Reversible Terminators [Publisher: National Academy of Sciences]. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(52), 19635–19640.

Karnani, M. M., Jackson, J., Ayzenshtat, I., Tucciarone, J., Manoocheri, K., Snider, W. G., & Yuste, R. (2016). Cooperative Subnetworks of Molecularly Similar Interneurons in Mouse Neocortex [Publisher: Elsevier]. *Neuron*, *90*(1), 86–100. https://doi.org/10.1016/j.neuron.2016.02.037

Katz, L. C., & Dalva, M. B. (1994). Scanning laser photostimulation: A new approach for analyzing brain circuits. *Journal of Neuroscience Methods*, *54*(2), 205–218. https://doi.org/10.1016/0165-0270(94)90194-5

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., & Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells [Publisher: Nature Publishing Group]. *Nature Methods*, *10*(9), 857–860. https://doi.org/10.1038/nmeth.2563

Kebschull, J. M., Garcia da Silva, P., Reid, A. P., Peikon, I. D., Albeanu, D. F., & Zador, A. M. (2016). High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA [MAPseq]. *Neuron*, *91*(5), 975–987. https://doi.org/10.1016/j.neuron.2016.07.036

Kebschull, J. M., Garcia da Silva, P., & Zador, A. M. (2016). A New Defective Helper RNA to Produce Recombinant Sindbis Virus that Infects Neurons but does not Propagate. *Frontiers in Neuroanatomy*, *10*. https://doi.org/10.3389/fnana.2016.00056

Kebschull, J. M., & Zador, A. M. (2018). Cellular barcoding: Lineage tracing, screening and beyond. *Nature Methods*, *15*(11), 871–879. https://doi.org/10.1038/s41592-018-0185-x

Kelly, R. M., & Strick, P. L. (2000). Rabies as a transneuronal tracer of circuits in the central nervous system. *Journal of Neuroscience Methods*, *103*(1), 63–71. https://doi.org/10.1016/s0165-0270(00)00296-x

Kelly, R. M., & Strick, P. L. (2003). Cerebellar Loops with Motor Cortex and Prefrontal Cortex of a Nonhuman Primate [Publisher: Society for Neuroscience Section: Behavioral/Systems/Cognitive]. *Journal of Neuroscience*, *23*(23), 8432–8444. https://doi.org/10.1523/JNEUROSCI.23-23-08432.2003

Kelly, R. M., & Strick, P. L. (2004, January). Macro-architecture of basal ganglia loops with the cerebral cortex: Use of rabies virus to reveal multisynaptic circuits. In *Progress in Brain Research* (pp. 447–459, Vol. 143). Elsevier. https://doi.org/10.1016/S0079-6123(03)43042-2

Kim, Y., Cheng, W., Cho, C.-S., Hwang, Y., Si, Y., Park, A., Schrank, M., Hsu, J.-E., Anacleto, A., Xi, J., Kim, M., Pedersen, E., Koues, O. I., Wilson, T., Lee, C., Jun, G., Kang, H. M., & Lee, J. H. (2024). Seq-Scope: Repurposing Illumina sequencing flow cells for high-resolution spatial transcriptomics [Publisher: Nature Publishing Group]. *Nature Protocols*, 1–47. https://doi.org/10.1038/s41596-024-01065-0

Klein, S., Staring, M., Murphy, K., Viergever, M., & Pluim, J. (2010). Elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, *29*(1), 196–205. https://doi.org/10.1109/TMI.2009.2035616

Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S. B., & Mrsic-Flogel, T. D. (2013). The emergence of functional microcircuits in visual cortex. *Nature*, *496*(7443), 96–100. https://doi.org/10.1038/nature12015

Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjöström, P. J., & Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, *473*(7345), 87–91. https://doi.org/10.1038/nature09880

Kobiler, O., Lipman, Y., Therkelsen, K., Daubechies, I., & Enquist, L. W. (2010). Herpesviruses carrying a Brainbow cassette reveal replication and expression of limited numbers of incoming genomes [Publisher: Nature Publishing Group]. *Nature Communications*, *1*(1), 146. https://doi.org/10.1038/ncomms1145

Kolb, I., Landry, C. R., Yip, M. C., Lewallen, C. F., Stoy, W. A., Lee, J., Felouzis, A., Yang, B., Boyden, E. S., Rozell, C. J., & Forest, C. R. (2019). PatcherBot: A single-cell electrophysiology robot for adherent cells and brain slices. *Journal of Neural Engineering*, *16*(4), 046003. https://doi.org/10.1088/1741-2552/ab1834

Koos, K., Oláh, G., Balassa, T., Mihut, N., Rózsa, M., Ozsvár, A., Tasnadi, E., Barzó, P., Faragó, N., Puskás, L., Molnár, G., Molnár, J., Tamás, G., & Horvath, P. (2021). Automatic deep learning-driven label-free image-guided patch clamp system [Publisher: Nature Publishing Group]. *Nature Communications*, *12*(1), 936. https://doi.org/10.1038/s41467-021-21291-4

Krzywkowski, T., & Nilsson, M. (2017). Fidelity of RNA templated end-joining by chlorella virus DNA ligase and a novel iLock assay with improved direct RNA detection accuracy. *Nucleic Acids Research*, *45*(18), e161. https://doi.org/10.1093/nar/gkx708

Kuan, A. T., Phelps, J. S., Thomas, L. A., Nguyen, T. M., Han, J., Chen, C.-L., Azevedo, A. W., Tuthill, J. C., Funke, J., Cloetens, P., Pacureanu, A., & Lee, W.-C. A. (2020). Dense neuronal reconstruction through X-ray holographic nano-tomography. *Nature Neuroscience*, *23*(12), 1637–1643. https://doi.org/10.1038/s41593-020-0704-9

Kucera, P., Dolivo, M., Coulon, P., & Flamand, A. (1985). Pathways of the early propagation of virulent and avirulent rabies strains from the eye to the brain. *Journal of Virology*, *55*(1), 158–162. https://doi.org/10.1128/JVI.55.1.158-162.1985

Kuramoto, E. (2019). Method for labeling and reconstruction of single neurons using Sindbis virus vectors. *Journal of Chemical Neuroanatomy*, *100*, 101648. https://doi.org/10.1016/j.jchemneu.2019.05.002

Lahaye, X., Vidy, A., Pomier, C., Obiang, L., Harper, F., Gaudin, Y., & Blondel, D. (2009). Functional Characterization of Negri Bodies (NBs) in Rabies Virus-Infected Cells: Evidence that NBs Are Sites of Viral Transcription and Replication [Publisher: American Society for Microbiology]. *Journal of Virology*, *83*(16), 7948–7958. https://doi.org/10.1128/jvi.00554-09

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., … Zhang, K. (2016). Neuronal subtypes and diversity revealed by

single-nucleus RNA sequencing of the human brain [Publisher: American Association for the Advancement of Science]. *Science*. https://doi.org/10.1126/science.aaf1204

Lappalainen, J. K., Tschopp, F. D., Prakhya, S., McGill, M., Nern, A., Shinomiya, K., Takemura, S.-Y., Gruntman, E., Macke, J. H., & Turaga, S. C. (2024). Connectome-constrained networks predict neural activity across the fly visual system. *Nature*, *634*(8036), 1132–1140. https://doi.org/10.1038/s41586-024-07939-3

Lavin, T. K., Jin, L., Lea, N. E., & Wickersham, I. R. (2020). Monosynaptic Tracing Success Depends Critically on Helper Virus Concentrations. *Frontiers in Synaptic Neuroscience*, *12*, 6. https://doi.org/10.3389/fnsyn.2020.00006

Lawson, N. D., Stillman, E. A., Whitt, M. A., & Rose, J. K. (1995). Recombinant vesicular stomatitis viruses from DNA. [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *92*(10), 4477–4481. https://doi.org/10.1073/pnas.92.10.4477

Lee, B. R., Budzillo, A., Hadley, K., Miller, J. A., Jarsky, T., Baker, K., Hill, D., Kim, L., Mann, R., Ng, L., Oldre, A., Rajanbabu, R., Trinh, J., Vargas, S., Braun, T., Dalley, R. A., Gouwens, N. W., Kalmbach, B. E., Kim, T. K., … Berg, J. (2021). Scaled, high fidelity electrophysiological, morphological, and transcriptomic cell characterization (S. B. Nelson, L. Chen, S. B. Nelson, & C. J. McBain, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *10*, e65482. https://doi.org/10.7554/eLife.65482

Lee, B. R., Dalley, R., Miller, J. A., Chartrand, T., Close, J., Mann, R., Mukora, A., Ng, L., Alfiler, L., Baker, K., Bertagnolli, D., Brouner, K., Casper, T., Csajbok, E., Donadio, N., Driessens, S. L., Egdorf, T., Enstrom, R., Galakhova, A. A., … Ting, J. T. (2023). Signature morphoelectric properties of diverse GABAergic interneurons in the human neocortex [Publisher: American Association for the Advancement of Science]. *Science*, *382*(6667), eadf6484. https://doi.org/10.1126/science.adf6484

Lee, H., Marco Salas, S., Gyllborg, D., & Nilsson, M. (2022). Direct RNA targeted in situ sequencing for transcriptomic profiling in tissue [Publisher: Nature Publishing Group]. *Scientific Reports*, *12*(1), 7976. https://doi.org/10.1038/s41598-022-11534-9

Lee, S., Kruglikov, I., Huang, Z. J., Fishell, G., & Rudy, B. (2013). A disinhibitory circuit mediates motor integration in the somatosensory cortex [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *16*(11), 1662–1670. https://doi.org/10.1038/nn.3544

Lee, W.-C. A., Bonin, V., Reed, M., Graham, B. J., Hood, G., Glattfelder, K., & Reid, R. C. (2016). Anatomy and function of an excitatory network in the visual cortex. *Nature*, *532*(7599), 370–374. https://doi.org/10.1038/nature17192

Lefort, S., & Petersen, C. C. H. (2017). Layer-Dependent Short-Term Synaptic Plasticity Between Excitatory Neurons in the C2 Barrel Column of Mouse Primary Somatosensory Cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(7), 3869–3878. https://doi.org/10.1093/cercor/bhx094

Lefort, S., Tomm, C., Sarria, J.-C. F., & Petersen, C. C. H. (2009). The Excitatory Neuronal Network of the C2 Barrel Column in Mouse Primary Somatosensory Cortex [Publisher: Elsevier]. *Neuron*, *61*(2), 301–316. https://doi.org/10.1016/j.neuron.2008.12.020

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T.-M., Chi Chin, M., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., … Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain [Number: 7124 Publisher: Nature Publishing Group]. *Nature*, *445*(7124), 168–176. https://doi.org/10.1038/nature05453

Lépine, P. (1938). On the Evolution of Fixed Strains of Rabies Virus [Publisher: Cambridge University Press]. *The Journal of Hygiene*, *38*(2), 180–184.

Levy, R. B., & Reyes, A. D. (2012). Spatial Profile of Excitatory and Inhibitory Synaptic Connectivity in Mouse Primary Auditory Cortex. *Journal of Neuroscience*, *32*(16), 5609–5619. https://doi.org/10.1523/JNEUROSCI.5158-11.2012

Levy, S. F., Blundell, J. R., Venkataram, S., Petrov, D. A., Fisher, D. S., & Sherlock, G. (2015). Quantitative evolutionary dynamics using high-resolution lineage tracking [Publisher: Nature Publishing Group]. *Nature*, *519*(7542), 181–186. https://doi.org/10.1038/nature14279

Li, C.-L., Li, K.-C., Wu, D., Chen, Y., Luo, H., Zhao, J.-R., Wang, S.-S., Sun, M.-M., Lu, Y.-J., Zhong, Y.-Q., Hu, X.-Y., Hou, R., Zhou, B.-B., Bao, L., Xiao, H.-S., & Zhang, X. (2016). Somatosensory neuron types identified by high-coverage single-cell RNA-sequencing and functional heterogeneity [Publisher: Nature Publishing Group]. *Cell Research*, *26*(1), 83–102. https://doi.org/10.1038/cr.2015.149

Li, N., Chen, T.-W., Guo, Z. V., Gerfen, C. R., & Svoboda, K. (2015). A motor cortex circuit for motor planning and movement. *Nature*, *519*(7541), 51–56. https://doi.org/10.1038/nature14178

Li, Y., Stam, F. J., Aimone, J. B., Goulding, M., Callaway, E. M., & Gage, F. H. (2013). Molecular layer perforant path-associated cells contribute to feed-forward inhibition in the adult dentate gyrus [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *110*(22), 9106–9111. https://doi.org/10.1073/pnas.1306912110

Li, Y., Walker, L. A., Zhao, Y., Edwards, E. M., Michki, N. S., Cheng, H. P. J., Ghazzi, M., Chen, T. Y., Chen, M., Roossien, D. H., & Cai, D. (2021). Bitbow Enables Highly Efficient Neuronal Lineage Tracing and Morphology Reconstruction in Single Drosophila Brains. *Frontiers in Neural Circuits*, *15*, 732183. https://doi.org/10.3389/fncir.2021.732183

Liang, L., Zhou, P., Ye, C., Yang, Q., & Ma, T. (2023). Spatial–temporal patterns of brain disconnectome in Alzheimer's disease. *Human Brain Mapping*, *44*(11), 4272–4286. https://doi.org/10.1002/hbm.26344

Lin, Z., Ophir, O., Trimbuch, T., Brokowski, B., Tibi, M., Pavletsov, P., Schmidt, H., Rosenmund, C., Hochgerner, H., & Zeisel, A. (2024, April). Brain-wide monosynaptic connectivity mapping with ROInet-seq [Pages: 2024.04.04.588058 Section: New Results]. https://doi.org/10.1101/2024.04.04.588058

Liu, J., Tran, V., Vemuri, V. N. P., Byrne, A., Borja, M., Kim, Y. J., Agarwal, S., Wang, R., Awayan, K., Murti, A., Taychameekiatchai, A., Wang, B., Emanuel, G., He, J., Haliburton, J., Oliveira Pisco, A., & Neff, N. F. (2023). Concordance of MERFISH spatial transcriptomics with bulk and single-cell RNA sequencing. *Life Science Alliance*, *6*(1), e202201701. https://doi.org/10.26508/lsa.202201701

Liu, S., Punthambaker, S., Iyer, E. P. R., Ferrante, T., Goodwin, D., Fürth, D., Pawlowski, A. C., Jindal, K., Tam, J. M., Mifflin, L., Alon, S., Sinha, A., Wassie, A. T., Chen, F., Cheng, A., Willocq, V., Meyer, K., Ling, K.-H., Camplisson, C. K., … Church, G. M. (2021). Barcoded oligonucleotides ligated on RNA amplified for multiplexed and parallel *in situ* analyses. *Nucleic Acids Research*, *49*(10), e58–e58. https://doi.org/10.1093/nar/gkab120

Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, 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*(6), 1665–1681.e18. https://doi.org/10.1016/j.cell.2020.10.026

Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., & Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector [Publisher: Nature Publishing Group]. *Scientific Reports*, *7*(1), 2193. https://doi.org/10.1038/s41598-017-02460-2

Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R., & Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, *450*(7166), 56–62. https://doi.org/10.1038/nature06293

Lohman, G. J. S., Zhang, Y., Zhelkovsky, A. M., Cantor, E. J., & Evans, T. C., Jr. (2014). Efficient DNA ligation in DNA–RNA hybrid helices by Chlorella virus DNA ligase. *Nucleic Acids Research*, *42*(3), 1831–1844. https://doi.org/10.1093/nar/gkt1032

Lu, X., Wu, Y., Schalek, R. L., Meirovitch, Y., Berger, D. R., & Lichtman, J. W. (2023). A Scalable Staining Strategy for Whole-Brain Connectomics. *bioRxiv: The Preprint Server for Biology*, 2023.09.26.558265. https://doi.org/10.1101/2023.09.26.558265

Lucić, V., Yang, T., Schweikert, G., Förster, F., & Baumeister, W. (2005). Morphological characterization of molecular complexes present in the synaptic cleft. *Structure (London, England: 1993)*, *13*(3), 423–434. https://doi.org/10.1016/j.str.2005.02.005

Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: A tutorial [Publisher: John Wiley & Sons, Ltd]. *Molecular Systems Biology*, *15*(6), e8746. https://doi.org/10.15252/msb.20188746

Lundh, B., Löve, A., Kristensson, K., & Norrby, E. (1988). Non-Lethal Infection of Aminergic Reticular Core Neurons: Age-Dependent Spread of ts Mutant Vesicular Stomatitis Virus from the Nose. *Journal of Neuropathology & Experimental Neurology*, *47*(5), 497–506. https://doi.org/10.1097/00005072-198809000-00001

Lyman, M. G., Feierbach, B., Curanovic, D., Bisher, M., & Enquist, L. W. (2007). Pseudorabies Virus Us9 Directs Axonal Sorting of Viral Capsids [Publisher: American Society for Microbiology]. *Journal of Virology*, *81*(20), 11363–11371. https://doi.org/10.1128/jvi.01281-07

Ma, Y., Hu, H., & Agmon, A. (2012). Short-Term Plasticity of Unitary Inhibitory-to-Inhibitory Synapses Depends on the Presynaptic Interneuron Subtype [Publisher: Society for Neuroscience Section: Brief Communications]. *Journal of Neuroscience*, *32*(3), 983–988. https://doi.org/10.1523/JNEUROSCI.5007-11.2012

Mahalingam, G., Torres, R., Kapner, D., Trautman, E. T., Fliss, T., Seshamani, S., Perlman, E., Young, R., Kinn, S., Buchanan, J., Takeno, M. M., Yin, W., Bumbarger, D. J., Gwinn, R. P., Nyhus, J., Lein, E., Smith, S. J., Reid, R. C., Khairy, K. A., … Macarico da Costa, N. (2022). A scalable and modular automated pipeline for stitching of large electron microscopy datasets (A. Cardona, R. L. Calabrese, A. Cardona, I. Arganda-Carreras, & C. Tischer, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *11*, e76534. https://doi.org/10.7554/eLife.76534

Mandino, F., Shen, X., Desrosiers-Grégoire, G., O'Connor, D., Mukherjee, B., Owens, A., Qu, A., Onofrey, J., Papademetris, X., Chakravarty, M. M., Strittmatter, S. M., & Lake, E. M. R. (2024). Aging-dependent loss of functional connectivity in a mouse model of Alzheimer's disease and reversal by mGluR5 modulator [Publisher: Nature Publishing Group]. *Molecular Psychiatry*, 1–16. https://doi.org/10.1038/s41380-024-02779-z

Mangini, N. J., & Pearlman, A. L. (1980). Laminar distribution of receptive field properties in the primary visual cortex of the mouse [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.901930114]. *Journal of Comparative Neurology*, *193*(1), 203–222. https://doi.org/10.1002/cne.901930114

Marshel, J. H., Mori, T., Nielsen, K. J., & Callaway, E. M. (2010). Targeting Single Neuronal Networks for Gene Expression and Cell Labeling In Vivo. *Neuron*, *67*(4), 562–574. https://doi.org/10.1016/j.neuron.2010.08.001

Masaki, Y., Yamaguchi, M., Takeuchi, R. F., & Osakada, F. (2022). Monosynaptic rabies virus tracing from projection-targeted single neurons. *Neuroscience Research*, *178*, 20–32. https://doi.org/10.1016/j.neures.2022.01.007

Mathis, V. P., Williams, M., Fillinger, C., & Kenny, P. J. (2021). Networks of habenula-projecting cortical neurons regulate cocaine seeking [Publisher: American Association for the Advancement of Science]. *Science Advances*, *7*(45), eabj2225. https://doi.org/10.1126/sciadv.abj2225

Mayerich, D., Abbott, L., & McCORMICK, B. (2008). Knife-edge scanning microscopy for imaging and reconstruction of three-dimensional anatomical structures of the mouse brain. *Journal of Microscopy*, *231*(1), 134–143. https://doi.org/10.1111/j.1365-2818.2008.02024.x

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 [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *24*(3), 425–436. https://doi.org/10.1038/s41593-020-00787-0

Meurens, F., Schynts, F., Keil, G. M., Muylkens, B., Vanderplasschen, A., Gallego, P., & Thiry, E. (2004). Superinfection prevents recombination of the alphaherpesvirus bovine herpesvirus 1. *Journal of Virology*, *78*(8), 3872–3879. https://doi.org/10.1128/jvi.78.8.3872-3879.2004

MICrONS Consortium, Bae, J. A., Baptiste, M., Bodor, A. L., Brittain, D., Buchanan, J., Bumbarger, D. J., Castro, M. A., Celii, B., Cobos, E., Collman, F., da Costa, N. M., Dorkenwald, S., Elabbady, L., Fahey, P. G., Fliss, T., Froudarakis, E., Gager, J., Gamlin, C., … Yu, S.-c. (2021, July). *Functional connectomics spanning multiple areas of mouse visual cortex* (preprint). Neuroscience. https://doi.org/10.1101/2021.07.28.454025

Miyamichi, K., Shlomai-Fuchs, Y., Shu, M., Weissbourd, B. C., Luo, L., & Mizrahi, A. (2013). Dissecting Local Circuits: Parvalbumin Interneurons Underlie Broad Feedback Control of Olfactory Bulb Output [Publisher: Elsevier]. *Neuron*, *80*(5), 1232–1245. https://doi.org/10.1016/j.neuron.2013.08.027

Miyamichi, K., Amat, F., Moussavi, F., Wang, C., Wickersham, I., Wall, N. R., Taniguchi, H., Tasic, B., Huang, Z. J., He, Z., Callaway, E. M., Horowitz, M. A., & Luo, L. (2011). Cortical representations of olfactory input by trans-synaptic tracing. *Nature*, *472*(7342), 191–196. https://doi.org/10.1038/nature09714

Mori, T., & Morimoto, K. (2014). Rabies virus glycoprotein variants display different patterns in rabies monosynaptic tracing [Publisher: Frontiers]. *Frontiers in Neuroanatomy*, *7*. https://doi.org/10.3389/fnana.2013.00047

Morimoto, K., Hooper, D. C., Carbaugh, H., Fu, Z. F., Koprowski, H., & Dietzschold, B. (1998). Rabies virus quasispecies: Implications for pathogenesis. *Proceedings of the National Academy of Sciences*, *95*(6), 3152–3156. https://doi.org/10.1073/pnas.95.6.3152

Muñoz-Castañeda, R., Zingg, B., Matho, K. S., Wang, Q., Chen, X., Foster, N. N., Narasimhan, A., Li, A., Hirokawa, K. E., Huo, B., Bannerjee, S., Korobkova, L., Park, C. S., Park, Y.-G., Bienkowski, M. S., Chon, U., Wheeler, D. W., Li, X., Wang, Y., … Dong, H.-W. (2020, October). *Cellular Anatomy of the Mouse Primary Motor Cortex* (preprint). Neuroscience. https://doi.org/10.1101/2020.10.02.323154

Murakami, T. C., Mano, T., Saikawa, S., Horiguchi, S. A., Shigeta, D., Baba, K., Sekiya, H., Shimizu, Y., Tanaka, K. F., Kiyonari, H., Iino, M., Mochizuki, H., Tainaka, K., & Ueda, H. R. (2018). A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *21*(4), 625–637. https://doi.org/10.1038/s41593-018-0109-1

Murray, K. T., Pedersen, A. F., Mohacsi, I., Detlefs, C., Morgan, A. J., Prasciolu, M., Yildirim, C., Simons, H., Jakobsen, A. C., Chapman, H. N., Poulsen, H. F., & Bajt, S. (2019). Multilayer Laue lenses at high X-ray energies: Performance and applications. *Optics Express*, *27*(5), 7120–7138. https://doi.org/10.1364/OE.27.007120

Naik, S. H., Perié, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R. J., & Schumacher, T. N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors [Publisher: Nature Publishing Group]. *Nature*, *496*(7444), 229–232. https://doi.org/10.1038/nature12013

Negri.A. (1903). *Sull'eziologia Della Rabbia – La Diagnosi Della Rabbia In Base Ai Nuovi Reperti*.

Nehar-Belaid, D., Hong, S., Marches, R., Chen, G., Bolisetty, M., Baisch, J., Walters, L., Punaro, M., Rossi, R. J., Chung, C.-H., Huynh, R. P., Singh, P., Flynn, W. F., Tabanor-Gayle, J.-A., Kuchipudi, N., Mejias, A., Collet, M. A., Lucido, A. L., Palucka, K., … Banchereau, J. F. (2020). Mapping systemic lupus erythematosus heterogeneity at the single-cell level [Publisher: Nature Publishing Group]. *Nature Immunology*, *21*(9), 1094–1106. https://doi.org/10.1038/s41590-020-0743-0

Ohara, S. (2009). Dual transneuronal tracing in the rat entorhinal-hippocampal circuit by intracerebral injection of recombinant rabies virus vectors. *Frontiers in Neuroanatomy*, *3*. https://doi.org/10.3389/neuro.05.001.2009

Osakada, F., & Callaway, E. M. (2013). Design and generation of recombinant rabies virus vectors [Publisher: Nature Publishing Group]. *Nature Protocols*, *8*(8), 1583–1601. https://doi.org/10.1038/nprot.2013.094

Oyibo, H., Cao, C., Ferrante, D. D., Zhan, H., Koulakov, A., Enquist, L., Dubnau, J., & Zador, A. (2018, January). *A computational framework for converting high-throughput DNA sequencing data into neural circuit connectivity* (preprint). Neuroscience.

Padfield, D. (2012). Masked Object Registration in the Fourier Domain [Conference Name: IEEE Transactions on Image Processing]. *IEEE Transactions on Image Processing*, *21*(5), 2706–2718. https://doi.org/10.1109/TIP.2011.2181402

Pasteur, L. (1884). Nouvelle communcation sur la rage [Publisher: Persée - Portail des revues scientifiques en SHS]. *Recueil de Médecine Vétérinaire*, *61*, 129–159. https://doi.org/10.3406/rmv.1884.1781

Pati, Y., Rezaiifar, R., & Krishnaprasad, P. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition [ISSN: 1058-6393]. *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, 40–44 vol.1. https://doi.org/10.1109/ACSSC.1993.342465

Patiño, M., Lagos, W. N., Patne, N. S., Tasic, B., Zeng, H., & Callaway, E. M. (2022). Single-cell transcriptomic classification of rabies-infected cortical neurons [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *119*(22), e2203677119. https://doi.org/10.1073/pnas.2203677119

Patiño, M., Rossa, M. A., Lagos, W. N., Patne, N. S., & Callaway, E. M. (2024). Transcriptomic cell-type specificity of local cortical circuits [Publisher: Elsevier]. *Neuron*, *112*(23), 3851–3866.e4. https://doi.org/10.1016/j.neuron.2024.09.003

Pattnaik, A. K., Ball, L. A., LeGrone, A. W., & Wertz, G. W. (1992). Infectious defective interfering particles of VSV from transcripts of a cDNA clone. *Cell*, *69*(6), 1011–1020. https://doi.org/10.1016/0092-8674(92)90619-n

Peng, Y., Mittermaier, F. X., Planert, H., Schneider, U. C., Alle, H., & Geiger, J. R. P. (2019). High-throughput microcircuit analysis of individual human brains through next-generation multineuron patch-clamp (J. Huguenard, E. Marder, & T. Jarsky, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *8*, e48178. https://doi.org/10.7554/eLife.48178

Peng, Y., Bjelde, A., Aceituno, P. V., Mittermaier, F. X., Planert, H., Grosser, S., Onken, J., Faust, K., Kalbhenn, T., Simon, M., Radbruch, H., Fidzinski, P., Schmitz, D., Alle, H., Holtkamp, M., Vida, I., Grewe, B. F., & Geiger, J. R. P. (2024). Directed and acyclic synaptic connectivity in the human layer 2-3 cortical microcircuit. *Science (New York, N.Y.)*, *384*(6693), 338–343. https://doi.org/10.1126/science.adg8828

Perin, R., Berger, T. K., & Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences*, *108*(13), 5419–5424. https://doi.org/10.1073/pnas.1016051108

Petreanu, L., Huber, D., Sobczyk, A., & Svoboda, K. (2007). Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neuroscience*, *10*(5), 663–668. https://doi.org/10.1038/nn1891

Petreanu, L., Mao, T., Sternson, S. M., & Svoboda, K. (2009). The subcellular organization of neocortical excitatory connections [Publisher: Nature Publishing Group]. *Nature*, *457*(7233), 1142–1145. https://doi.org/10.1038/nature07709

Petukhov, V., Xu, R. J., Soldatov, R. A., Cadinu, P., Khodosevich, K., Moffitt, J. R., & Kharchenko, P. V. (2022). Cell segmentation in imaging-based spatial transcriptomics [Publisher: Nature Publishing Group]. *Nature Biotechnology*, *40*(3), 345–354. https://doi.org/10.1038/s41587-021-01044-w

Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., & Scanziani, M. (2013). Inhibition of inhibition in visual cortex: The logic of connections between molecularly distinct interneurons. *Nature Neuroscience*, *16*(8), 1068–1076. https://doi.org/10.1038/nn.3446

Phelps, J. S., Hildebrand, D. G. C., Graham, B. J., Kuan, A. T., Thomas, L. A., Nguyen, T. M., Buhmann, J., Azevedo, A. W., Sustar, A., Agrawal, S., Liu, M., Shanny, B. L., Funke, J., Tuthill, J. C., & Lee, W.-C. A. (2021). Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy [Publisher: Elsevier]. *Cell*, *184*(3), 759–774.e18. https://doi.org/10.1016/j.cell.2020.12.013

Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., & Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control [Publisher: Nature Publishing Group]. *Nature*, *503*(7477), 521–524. https://doi.org/10.1038/nature12676

Pietilä, R., Gaudio, F. D., He, L., Vázquez-Liébanas, E., Vanlandewijck, M., Muhl, L., Mocci, G., Bjørnholm, K. D., Lindblad, C., Fletcher-Sandersjöö, A., Svensson, M., Thelin, E. P., Liu, J., Voorden, A. J. v., Torres, M., Antila, S., Xin, L., Karlström, H., Storm-Mathisen, J., … Betsholtz, C. (2023). Molecular anatomy of adult mouse leptomeninges [Publisher: Elsevier]. *Neuron*, *111*(23), 3745–3764.e7. https://doi.org/10.1016/j.neuron.2023.09.002

Pinkard, H., Stuurman, N., Ivanov, I. E., Anthony, N. M., Ouyang, W., Li, B., Yang, B., Tsuchida, M. A., Chhun, B., Zhang, G., Mei, R., Anderson, M., Shepherd, D. P., Hunt-Isaak, I., Dunn, R. L., Jahr, W., Kato, S., Royer, L. A., Thiagarajah, J. R., … Waller, L. (2021). Pycro-Manager: Open-source software for customized and reproducible microscope control [Publisher: Nature Publishing Group]. *Nature Methods*, *18*(3), 226–228. https://doi.org/10.1038/s41592-021-01087-6

Potratz, M., Zaeck, L. M., Weigel, C., Klein, A., Freuling, C. M., Müller, T., & Finke, S. (2020). Neuroglia infection by rabies virus after anterograde virus spread in peripheral neurons. *Acta Neuropathologica Communications*, *8*(1), 199. https://doi.org/10.1186/s40478-020-01074-6

Poulin, J.-F., Zou, J., Drouin-Ouellet, J., Kim, K.-Y. A., Cicchetti, F., & Awatramani, R. B. (2014). Defining Midbrain Dopaminergic Neuron Diversity by Single-Cell Gene Expression Profiling [Publisher: Elsevier]. *Cell Reports*, *9*(3), 930–943. https://doi.org/10.1016/j.celrep.2014.10.008

Prabhakar, S., Lule, S., da Hora, C. C., Breakefield, X. O., & Cheah, P. S. (2021). AAV9 transduction mediated by systemic delivery of vector via retro-orbital injection in newborn, neonatal and juvenile mice. *Experimental Animals*, *70*(4), 450–458. https://doi.org/10.1538/expanim.20-0186

Qian, X., Harris, K. D., Hauling, T., Nicoloutsopoulos, D., Muñoz-Manchado, A. B., Skene, N., Hjerling-Leffler, J., & Nilsson, M. (2020). Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nature Methods*, *17*(1), 101–106. https://doi.org/10.1038/s41592-019-0631-4

Radecke, F., Spielhofer, P., Schneider, H., Kaelin, K., Huber, M., Dötsch, C., Christiansen, G., & Billeter, M. A. (1995). Rescue of measles viruses from cloned DNA. *The EMBO Journal*, *14*(23), 5773–5784.

Ragan, T., Kadiri, L. R., Venkataraju, K. U., Bahlmann, K., Sutin, J., Taranda, J., Arganda-Carreras, I., Kim, Y., Seung, H. S., & Osten, P. (2012). Serial two-photon tomography: An automated method for ex-vivo mouse brain imaging. *Nature methods*, *9*(3), 255–258. https://doi.org/10.1038/nmeth.1854

Rancz, E. A., Franks, K. M., Schwarz, M. K., Pichler, B., Schaefer, A. T., & Margrie, T. W. (2011). Transfection via whole-cell recording in vivo: Bridging single-cell physiology, genetics and connectomics. *Nature Neuroscience*, *14*(4), 527–532. https://doi.org/10.1038/nn.2765

Rasalingam, P., Rossiter, J. P., & Jackson, A. C. (2005). Recombinant rabies virus vaccine strain SAD-L16 inoculated intracerebrally in young mice produces a severe encephalitis with extensive neuronal apoptosis. *Canadian Journal of Veterinary Research*, *69*(2), 100–105.

Ray, N. B., Power, C., Lynch, W. P., Ewalt, L. C., & Lodmell, D. L. (1997). Rabies viruses infect primary cultures of murine, feline, and human microglia and astrocytes. *Archives of Virology*, *142*(5), 1011–1019. https://doi.org/10.1007/s007050050136

Reardon, T. R., Murray, A. J., Turi, G. F., Wirblich, C., Croce, K. R., Schnell, M. J., Jessell, T. M., & Losonczy, A. (2016). Rabies Virus CVS-N2c G Strain Enhances Retrograde Synaptic Transfer and Neuronal Viability [N2c]. *Neuron*, *89*(4), 711–724. https://doi.org/10.1016/j.neuron.2016.01.004

Remaut, K., Sanders, N. N., Fayazpour, F., Demeester, J., & De Smedt, S. C. (2006). Influence of plasmid DNA topology on the transfection properties of DOTAP/DOPE lipoplexes. *Journal of Controlled Release*, *115*(3), 335–343. https://doi.org/10.1016/j.jconrel.2006.08.009

Reyes, A., Lujan, R., Rozov, A., Burnashev, N., Somogyi, P., & Sakmann, B. (1998). Target-cell-specific facilitation and depression in neocortical circuits [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *1*(4), 279–285. https://doi.org/10.1038/1092

Roberts, A., & Rose, J. K. (1998). Recovery of negative-strand RNA viruses from plasmid DNAs: A positive approach revitalizes a negative field. *Virology*, *247*(1), 1–6. https://doi.org/10.1006/viro.1998.9250

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, 6.

Rompani, S. B., Müllner, F. E., Wanner, A., Zhang, C., Roth, C. N., Yonehara, K., & Roska, B. (2017). Different Modes of Visual Integration in the Lateral Geniculate Nucleus Revealed by Single-Cell-Initiated Transsynaptic Tracing. *Neuron*, *93*(4), 767–776.e6. https://doi.org/10.1016/j.neuron.2017.01.028

Ros, I. G., Omoto, J. J., & Dickinson, M. H. (2024). Descending control and regulation of spontaneous flight turns in Drosophila. *Current biology: CB*, *34*(3), 531–540.e5. https://doi.org/10.1016/j.cub.2023.12.047

Rossi, L. F., Harris, K. D., & Carandini, M. (2020). Spatial connectivity matches direction selectivity in visual cortex. *Nature*, *588*(7839), 648–652. https://doi.org/10.1038/s41586-020-2894-4

Ryan, K., Lu, Z., & Meinertzhagen, I. A. (2016). The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling (E. Marder, Ed.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *5*, e16962. https://doi.org/10.7554/eLife.16962

Sacramento, D., Badrane, H., Bourhy, H., & Tordo, N. (1992). Molecular epidemiology of rabies virus in France: Comparison with vaccine strains [Publisher: Microbiology Society,]. *Journal of General Virology*, *73*(5), 1149–1158. https://doi.org/10.1099/0022-1317-73-5-1149

Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral mutation rates. *Journal of Virology*, *84*(19), 9733–9748. https://doi.org/10.1128/JVI.00694-10

Sarkar, D., Kang, J., Wassie, A. T., Schroeder, M. E., Peng, Z., Tarr, T. B., Tang, A.-H., Niederst, E. D., Young, J. Z., Su, H., Park, D., Yin, P., Tsai, L.-H., Blanpied, T. A., & Boyden, E. S. (2022). Revealing nanostructures in brain tissue via protein decrowding by iterative expansion microscopy [Publisher: Nature Publishing Group]. *Nature Biomedical Engineering*, *6*(9), 1057–1073. https://doi.org/10.1038/s41551-022-00912-3

Sarkar, S. (2022). Design, Alignment, and Usage of Infinity-Corrected Microscope. In V. Nechyporuk-Zloy (Ed.), *Principles of Light Microscopy: From Basic to Advanced* (pp. 17–56). Springer International Publishing. https://doi.org/10.1007/978-3-031-04477-9_2

Saunders, A., Huang, K. W., Vondrak, C., Hughes, C., Smolyar, K., Sen, H., Philson, A. C., Nemesh, J., Wysoker, A., Kashin, S., Sabatini, B. L., & McCarroll, S. A. (2021, September). *Ascertaining cells' synaptic connections and RNA expression simultaneously with massively barcoded rabies virus libraries* (preprint). Neuroscience. https://doi.org/10.1101/2021.09.06.459177

Scala, F., Kobak, D., Bernabucci, M., Bernaerts, Y., Cadwell, C. R., Castro, J. R., Hartmanis, L., Jiang, X., Laturnus, S., Miranda, E., Mulherkar, S., Tan, Z. H., Yao, Z., Zeng, H., Sandberg, R., Berens, P., & Tolias, A. S. (2021). Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature*, *598*(7879), 144–150. https://doi.org/10.1038/s41586-020-2907-3

Schalek, R., Wilson, A., Lichtman, J., Josh, M., Kasthuri, N., Berger, D., Seung, S., Anger, P., Hayworth, K., & Aderhold, D. (2012). ATUM-based SEM for High-Speed Large-Volume Biological Reconstructions. *Microscopy and Microanalysis*, *18*(S2), 572–573. https://doi.org/10.1017/S1431927612004710

Schepers, K., Swart, E., van Heijst, J. W. J., Gerlach, C., Castrucci, M., Sie, D., Heimerikx, M., Velds, A., Kerkhoven, R. M., Arens, R., & Schumacher, T. N. M. (2008). Dissecting T cell lineage relationships by cellular barcoding. *The Journal of Experimental Medicine*, *205*(10), 2309–2318. https://doi.org/10.1084/jem.20072462

Schikorski, T., & Stevens, C. F. (1997). Quantitative Ultrastructural Analysis of Hippocampal Excitatory Synapses [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *17*(15), 5858–5867. https://doi.org/10.1523/JNEUROSCI.17-15-05858.1997

Schlegel, P., Yin, Y., Bates, A. S., Dorkenwald, S., Eichler, K., Brooks, P., Han, D. S., Gkantia, M., dos Santos, M., Munnelly, E. J., Badalamente, G., Serratosa Capdevila, L., Sane, V. A., Fragniere, A. M. C., Kiassat, L., Pleijzier, M. W., Stürner, T., Tamimi, I. F. M., Dunne, C. R., … Jefferis, G. S. X. E. (2024). Whole-brain annotation and multi-connectome cell typing of Drosophila [Publisher: Nature Publishing Group]. *Nature*, *634*(8032), 139–152. https://doi.org/10.1038/s41586-024-07686-5

Schmidt, M., Motta, A., Sievers, M., & Helmstaedter, M. (2024). RoboEM: Automated 3D flight tracing for synaptic-resolution connectomics [Publisher: Nature Publishing Group]. *Nature Methods*, *21*(5), 908–913. https://doi.org/10.1038/s41592-024-02226-5

Schneider, N., & Meier, M. (2017). Efficient in situ detection of mRNAs using the Chlorella virus DNA ligase for padlock probe ligation [Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab]. *RNA*, *23*(2), 250–256. https://doi.org/10.1261/rna.057836.116

Schnell, M., Mebatsion, T., & Conzelmann, K. (1994). Infectious rabies viruses from cloned cDNA. *The EMBO Journal*, *13*(18), 4195–4203. https://doi.org/10.1002/j.1460-2075.1994.tb06739.x

Seidler, B., Schmidt, A., Mayr, U., Nakhai, H., Schmid, R. M., Schneider, G., & Saur, D. (2008). A Cre-loxP-based mouse model for conditional somatic gene expression and knockdown in vivo by using avian retroviral vectors [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *105*(29), 10137–10142. https://doi.org/10.1073/pnas.0800487105

Sergey, G., Denis, K., Ava, H., Gediminas, G., Viola, O., L, K. O., H.p, L. R., O', B. M., Roger, P., C, W. J., & Alex, d. M. (2018, October). Oxygen plasma focused ion beam scanning electron microscopy for biological samples [Pages: 457820 Section: New Results]. https://doi.org/10.1101/457820

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. (2016). Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development*, *143*(15), 2862–2867. https://doi.org/10.1242/dev.138560

Shapson-Coe, A., Januszewski, M., Berger, D. R., Pope, A., Wu, Y., Blakely, T., Schalek, R. L., Li, P. H., Wang, S., Maitin-Shepard, J., Karlupia, N., Dorkenwald, S., Sjostedt, E., Leavitt, L., Lee, D., Troidl, J., Collman, F., Bailey, L., Fitzmaurice, A., … Lichtman, J. W. (2024). A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution [Publisher: American Association for the Advancement of Science]. *Science*, *384*(6696), eadk4858. https://doi.org/10.1126/science.adk4858

Shen, F. Y., Harrington, M. M., Walker, L. A., Cheng, H. P. J., Boyden, E. S., & Cai, D. (2020). Light microscopy based approach for mapping connectivity with molecular specificity. *Nature Communications*, *11*(1), 4632. https://doi.org/10.1038/s41467-020-18422-8

Shen, S., Bryant, K. D., Brown, S. M., Randell, S. H., & Asokan, A. (2011). Terminal N-Linked Galactose Is the Primary Receptor for Adeno-associated Virus 9 * [Publisher: Elsevier]. *Journal of Biological Chemistry*, *286*(15), 13532–13540. https://doi.org/10.1074/jbc.M110.210922

Shin, D., Urbanek, M. E., Larson, H. H., Moussa, A. J., Lee, K. Y., Baker, D. L., Standen-Bloom, E., Ramachandran, S., Bogdanoff, D., Cadwell, C. R., & Nowakowski, T. J. (2024, October). High-Complexity Barcoded Rabies Virus for Scalable Circuit Mapping Using Single-Cell and Single-Nucleus Sequencing [Pages: 2024.10.01.616167 Section: New Results]. https://doi.org/10.1101/2024.10.01.616167

Shiu, P. K., Sterne, G. R., Spiller, N., Franconville, R., Sandoval, A., Zhou, J., Simha, N., Kang, C. H., Yu, S., Kim, J. S., Dorkenwald, S., Matsliah, A., Schlegel, P., Yu, S.-C., McKellar, C. E., Sterling, A., Costa, M., Eichler, K., Bates, A. S., … Scott, K. (2024). A Drosophila computational brain model reveals sensorimotor processing. *Nature*, *634*(8032), 210–219. https://doi.org/10.1038/s41586-024-07763-9

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., & Chklovskii, D. B. (2005). Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits. *PLoS Biology*, *3*(3), e68. https://doi.org/10.1371/journal.pbio.0030068

Soto, G., Young, S., & Hama, K. (1994). Serial Section Electron Tomography: A Method for Three-Dimensional Reconstruction of Large Structures [Publisher: Academic Press]. *NeuroImage*, *1*(3), 230–243. https://doi.org/10.1006/nimg.1994.1008

Sountoulidis, A., Liontos, A., Nguyen, H. P., Firsova, A. B., Fysikopoulos, A., Qian, X., Seeger, W., Sundström, E., Nilsson, M., & Samakovlis, C. (2020). SCRINSHOT enables spatial mapping of cell states in tissue sections with single-cell resolution [Publisher: Public Library of Science]. *PLOS Biology*, *18*(11), e3000675. https://doi.org/10.1371/journal.pbio.3000675

Stepanyants, A., Martinez, L. M., Ferecskó, A. S., & Kisvárday, Z. F. (2009). The fractions of short- and long-range connections in the visual cortex. *Proceedings of the National Academy of Sciences*, *106*(9), 3555–3560. https://doi.org/10.1073/pnas.0810390106

Stepien, A. E., Tripodi, M., & Arber, S. (2010). Monosynaptic Rabies Virus Reveals Premotor Network Organization and Synaptic Specificity of Cholinergic Partition Cells [Publisher: Elsevier]. *Neuron*, *68*(3), 456–472. https://doi.org/10.1016/j.neuron.2010.10.019

Stickels, R. R., Murray, E., Kumar, P., Li, J., Marshall, J. L., Di Bella, D. J., Arlotta, P., Macosko, E. Z., & Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature Biotechnology*, *39*(3), 313–319. https://doi.org/10.1038/s41587-020-0739-1

Strack, A. M., Sawyer, W. B., Platt, K. B., & Loewy, A. D. (1989). CNS cell groups regulating the sympathetic outflow to adrenal gland as revealed by transneuronal cell body labeling with pseudorabies virus. *Brain Research*, *491*(2), 274–296. https://doi.org/10.1016/0006-8993(89)90063-2

Stringer, C., & Pachitariu, M. (2024). Cellpose3: One-click image restoration for improved cellular segmentation [Repository: Bioinformatics]. https://doi.org/10.1101/2024.02.10.579780

Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: A generalist algorithm for cellular segmentation [Publisher: Nature Publishing Group]. *Nature Methods*, *18*(1), 100–106. https://doi.org/10.1038/s41592-020-01018-x

Summerford, C., & Samulski, R. J. (1998). Membrane-Associated Heparan Sulfate Proteoglycan Is a Receptor for Adeno-Associated Virus Type 2 Virions [Publisher: American Society for Microbiology]. *Journal of Virology*, *72*(2), 1438–1445. https://doi.org/10.1128/jvi.72.2.1438-1445.1998

Sumser, A., Joesch, M., Jonas, P., & Ben-Simon, Y. (2022). Fast, high-throughput production of improved rabies viral vectors for specific, efficient and versatile transsynaptic retrograde labeling (R. Seal,

L. L. Colgin, & I. Wickersham, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *11*, e79848. https://doi.org/10.7554/eLife.79848

Sun, Y., Nguyen, A. Q., Nguyen, J. P., Le, L., Saur, D., Choi, J., Callaway, E. M., & Xu, X. (2014). Cell-Type-Specific Circuit Connectivity of Hippocampal CA1 Revealed through Cre-Dependent Rabies Tracing [Publisher: Elsevier]. *Cell Reports*, *7*(1), 269–280. https://doi.org/10.1016/j.celrep.2014.02.030

Sun, Y., Jin, S., Lin, X., Chen, L., Qiao, X., Jiang, L., Zhou, P., Johnston, K. G., Golshani, P., Nie, Q., Holmes, T. C., Nitz, D. A., & Xu, X. (2019). CA1-projecting subiculum neurons facilitate object-place learning. *Nature Neuroscience*, *22*(11), 1857–1870. https://doi.org/10.1038/s41593-019-0496-y

Sun, Y.-C., Chen, X., Fischer, S., Lu, S., Zhan, H., Gillis, J., & Zador, A. M. (2021). Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nature Neuroscience*, *24*(6), 873–885. https://doi.org/10.1038/s41593-021-00842-4

Tamás, G., Buhl, E. H., Lörincz, A., & Somogyi, P. (2000). Proximally targeted GABAergic synapses and gap junctions synchronize cortical interneurons [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *3*(4), 366–371. https://doi.org/10.1038/73936

Tang, M., Zhong, L., Rong, H., Li, K., Ye, M., Peng, J., & Ge, J. (2024). Efficient retinal ganglion cells transduction by retro-orbital venous sinus injection of AAV-PHP.eB in mature mice. *Experimental Eye Research*, *244*, 109931. https://doi.org/10.1016/j.exer.2024.109931

Tang, Y., Rampin, O., Giuliano, F., & Ugolini, G. (1999). Spinal and brain circuits to motoneurons of the bulbospongiosus muscle: Retrograde transneuronal tracing with rabies virus. *The Journal of Comparative Neurology*, *414*(2), 167–192.

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S. M., … Zeng, H. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience*, *19*(2), 335–346. https://doi.org/10.1038/nn.4216

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., … Zeng, H. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, *563*(7729), 72–78. https://doi.org/10.1038/s41586-018-0654-5

Tervo, D. G. R., Hwang, B.-Y., Viswanathan, S., Gaj, T., Lavzin, M., Ritola, K. D., Lindo, S., Michael, S., Kuleshova, E., Ojala, D., Huang, C.-C., Gerfen, C. R., Schiller, J., Dudman, J. T., Hantman, A. W., Looger, L. L., Schaffer, D. V., & Karpova, A. Y. (2016). A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons [Publisher: Elsevier]. *Neuron*, *92*(2), 372–382. https://doi.org/10.1016/j.neuron.2016.09.021

Thoulouze, M. I., Lafage, M., Schachner, M., Hartmann, U., Cremer, H., & Lafon, M. (1998). The neural cell adhesion molecule is a receptor for rabies virus. *Journal of Virology*, *72*(9), 7181–7190. https://doi.org/10.1128/JVI.72.9.7181-7190.1998

Tian, B., Zhou, M., Yang, Y., Yu, L., Luo, Z., Tian, D., Wang, K., Cui, M., Chen, H., Fu, Z. F., & Zhao, L. (2018). Lab-Attenuated Rabies Virus Causes Abortive Infection and Induces Cytokine Expression in Astrocytes by Activating Mitochondrial Antiviral-Signaling Protein Signaling Pathway. *Frontiers in Immunology*, *8*, 2011. https://doi.org/10.3389/fimmu.2017.02011

Tillberg, P. W., Chen, F., Piatkevich, K. D., Zhao, Y., Yu, C.-C., English, B. P., Gao, L., Martorell, A., Suk, H.-J., Yoshida, F., DeGennaro, E. M., Roossien, D. H., Gong, G., Seneviratne, U., Tannenbaum, S. R., Desimone, R., Cai, D., & Boyden, E. S. (2016). Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. *Nature Biotechnology*, *34*(9), 987–992. https://doi.org/10.1038/nbt.3625

Truckenbrodt, S., Maidorn, M., Crzan, D., Wildhagen, H., Kabatas, S., & Rizzoli, S. O. (2018). X10 expansion microscopy enables 25-nm resolution on conventional microscopes [Publisher: John Wiley & Sons, Ltd]. *EMBO reports*, *19*(9), e45836. https://doi.org/10.15252/embr.201845836

Tsiang, H., Koulakoff, A., Bizzini, B., & Berwald-Netter, Y. (1983). Neurotropism of rabies virus. An in vitro study. *Journal of Neuropathology and Experimental Neurology*, *42*(4), 439–452. https://doi.org/10.1097/00005072-198307000-00006

Tsoi, S. Y., Öncül, M., Svahn, E., Robertson, M., Bogdanowicz, Z., McClure, C., & Sürmeli, G. (2022). Telencephalic outputs from the medial entorhinal cortex are copied directly to the hippocampus (L. Giocomo, L. L. Colgin, & K. Allen, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *11*, e73162. https://doi.org/10.7554/eLife.73162

Tuffereau, C., Bénéjean, J., Blondel, D., Kieffer, B., & Flamand, A. (1998). Low-affinity nerve-growth factor receptor (P75NTR) can serve as a receptor for rabies virus. *The EMBO journal*, *17*(24), 7250–7259. https://doi.org/10.1093/emboj/17.24.7250

Tyson, A. L., Rousseau, C. V., Niedworok, C. J., Keshavarzi, S., Tsitoura, C., Cossell, L., Strom, M., & Margrie, T. W. (2021). A deep learning algorithm for 3D cell detection in whole mouse brain image datasets [Publisher: Public Library of Science]. *PLOS Computational Biology*, *17*(5), e1009074. https://doi.org/10.1371/journal.pcbi.1009074

Ugolini, G., Kuypers, H. G., & Simmons, A. (1987). Retrograde transneuronal transfer of herpes simplex virus type 1 (HSV 1) from motoneurones. *Brain Research*, *422*(2), 242–256. https://doi.org/10.1016/0006-8993(87)90931-0

Ugolini, G., Kuypers, H. G., & Strick, P. L. (1989). Transneuronal transfer of herpes virus from peripheral nerves to cortex and brainstem. *Science (New York, N.Y.)*, *243*(4887), 89–91. https://doi.org/10.1126/science.2536188

Ugolini, G. (1995). Specificity of rabies virus as a transneuronal tracer of motor networks: Transfer from hypoglossal motoneurons to connected second-order and higher order central nervous system cell groups [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.903560312]. *Journal of Comparative Neurology*, *356*(3), 457–480. https://doi.org/10.1002/cne.903560312

Ugolini, G. (2011). Rabies virus as a transneuronal tracer of neuronal connections. *Advances in Virus Research*, *79*, 165–202. https://doi.org/10.1016/B978-0-12-387040-7.00010-X

Uzonyi, A., Nir, R., & Schwartz, S. (2022). Cloning of DNA oligo pools for *in vitro* expression. *STAR Protocols*, *3*(1), 101103. https://doi.org/10.1016/j.xpro.2021.101103

van Brederode, J. F. M., Helliesen, M. K., & Hendrickson, A. E. (1991). Distribution of the calcium-binding proteins parvalbumin and calbindin-D28k in the sensorimotor cortex of the rat. *Neuroscience*, *44*(1), 157–171. https://doi.org/10.1016/0306-4522(91)90258-P

van den Pol, A. N., Ozduman, K., Wollmann, G., Ho, W. S. C., Simon, I., Yao, Y., Rose, J. K., & Ghosh, P. (2009). Viral strategies for studying the brain, including a replication-restricted self-amplifying delta-G vesicular stomatis virus that rapidly expresses transgenes in brain and can generate a multicolor golgi-like expression. *The Journal of Comparative Neurology*, *516*(6), 456–481. https://doi.org/10.1002/cne.22131

Vegué, M., Perin, R., & Roxin, A. (2017). On the Structure of Cortical Microcircuits Inferred from Small Sample Sizes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(35), 8498–8510. https://doi.org/10.1523/JNEUROSCI.0984-17.2017

Vélez-Fort, M., Rousseau, C. V., Niedworok, C. J., Wickersham, I. R., Rancz, E. A., Brown, A. P., Strom, M., & Margrie, T. W. (2014). The Stimulus Selectivity and Connectivity of Layer Six Principal Cells Reveals Cortical Microcircuits Underlying Visual Processing. *Neuron*, *83*(6), 1431–1443. https://doi.org/10.1016/j.neuron.2014.08.001

Verasztó, C., Jasek, S., Gühmann, M., Shahidi, R., Ueda, N., Beard, J. D., Mendes, S., Heinz, K., Bezares-Calderón, L. A., Williams, E., & Jékely, G. (2020, August). Whole-animal connectome and cell-

type complement of the three-segmented Platynereis dumerilii larva [Pages: 2020.08.21.260984 Section: New Results]. https://doi.org/10.1101/2020.08.21.260984

Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J. F., Gould, J., Griffin, G. K., Borg, Å., Ronaghi, M., Frisén, J., Lundeberg, J., Regev, A., & Ståhl, P. L. (2019). High-definition spatial transcriptomics for in situ tissue profiling. *Nature Methods*, *16*(10), 987–990. https://doi.org/10.1038/s41592-019-0548-y

Vladimirov, N., Voigt, F. F., Naert, T., Araujo, G. R., Cai, R., Reuss, A. M., Zhao, S., Schmid, P., Hildebrand, S., Schaettin, M., Groos, D., Mateos, J. M., Bethge, P., Yamamoto, T., Aerne, V., Roebroeck, A., Ertürk, A., Aguzzi, A., Ziegler, U., … Helmchen, F. (2024). Benchtop mesoSPIM: A next-generation open-source light-sheet microscope for cleared samples. *Nature Communications*, *15*(1), 2679. https://doi.org/10.1038/s41467-024-46770-2

Wadkins, T. S., Perrotta, A. T., Ferré-D'Amaré, A. R., Doudna, J. A., & Been, M. D. (1999). A nested double pseudoknot is required for self-cleavage activity of both the genomic and antigenomic hepatitis delta virus ribozymes. *RNA*, *5*(6), 720. https://doi.org/10.1017/s1355838299990209

Wall, N. R., Wickersham, I. R., Cetin, A., De La Parra, M., & Callaway, E. M. (2010). Monosynaptic circuit tracing in vivo through Cre-dependent targeting and complementation of modified rabies virus [Cre dependent TVA/G]. *Proceedings of the National Academy of Sciences*, *107*(50), 21848–21853. https://doi.org/10.1073/pnas.1011756107

Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J., & Luo, Y. (2012). RNAscope: A Novel in Situ RNA Analysis Platform for Formalin-Fixed, Paraffin-Embedded Tissues [Publisher: Elsevier]. *The Journal of Molecular Diagnostics*, *14*(1), 22–29. https://doi.org/10.1016/j.jmoldx.2011.08.002

Wang, H., Peca, J., Matsuzaki, M., Matsuzaki, K., Noguchi, J., Qiu, L., Wang, D., Zhang, F., Boyden, E., Deisseroth, K., Kasai, H., Hall, W. C., Feng, G., & Augustine, G. J. (2007). High-speed mapping of synaptic connectivity using photostimulation in Channelrhodopsin-2 transgenic mice [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *104*(19), 8143–8148. https://doi.org/10.1073/pnas.0700384104

Wang, J., Wang, Z., Liu, R., Shuai, L., Wang, X., Luo, J., Wang, C., Chen, W., Wang, X., Ge, J., He, X., Wen, Z., & Bu, Z. (2018). Metabotropic glutamate receptor subtype 2 is a cellular receptor for rabies virus [Publisher: Public Library of Science]. *PLOS Pathogens*, *14*(7), e1007189. https://doi.org/10.1371/journal.ppat.1007189

Wang, Q., Ding, S.-L., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer, B., Ho, A., Dolbeare, T., Blanchard, B., Dee, N., Wakeman, W., Hirokawa, K. E., Szafer, A., Sunkin, S. M., Oh, S. W., Bernard, A., … Ng, L. (2020). The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell*, *181*(4), 936–953.e20. https://doi.org/10.1016/j.cell.2020.04.007

Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., Nolan, G. P., Bava, F.-A., & Deisseroth, K. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states [Publisher: American Association for the Advancement of Science]. *Science*, *361*(6400), eaat5691. https://doi.org/10.1126/science.aat5691

Webb, J. M., Ma, M., Yin, C., Ptáček, L. J., & Fu, Y.-H. (2022). An excitatory peri-tegmental reticular nucleus circuit for wake maintenance [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *119*(34), e2203266119. https://doi.org/10.1073/pnas.2203266119

Weible, A. P., Schwarcz, L., Wickersham, I. R., DeBlander, L., Wu, H., Callaway, E. M., Seung, H. S., & Kentros, C. G. (2010). Transgenic Targeting of Recombinant Rabies Virus Reveals Monosynaptic Connectivity of Specific Neurons. *Journal of Neuroscience*, *30*(49), 16509–16513. https://doi.org/10.1523/JNEUROSCI.2442-10.2010

Weissbourd, B., Ren, J., DeLoach, K. E., Guenthner, C. J., Miyamichi, K., & Luo, L. (2014). Presynaptic Partners of Dorsal Raphe Serotonergic and GABAergic Neurons [Publisher: Elsevier]. *Neuron*, *83*(3), 645–662. https://doi.org/10.1016/j.neuron.2014.06.024

Wertz, A., Trenholm, S., Yonehara, K., Hillier, D., Raics, Z., Leinweber, M., Szalay, G., Ghanem, A., Keller, G., Rózsa, B., Conzelmann, K.-K., & Roska, B. (2015). Single-cell–initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science*, *349*(6243), 70–74. https://doi.org/10.1126/science.aab1687

Whelan, S. P., Ball, L. A., Barr, J. N., & Wertz, G. T. (1995). Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(18), 8388–8392. https://doi.org/10.1073/pnas.92.18.8388

Whitaker-Dowling, P., Ungner, J. S., Widnell, C. C., & Wilcox, D. K. (1983). Superinfect on exclusion by vesicular stomatitis virus. *Virology*, *131*(1), 137–143. https://doi.org/10.1016/0042-6822(83)90540-8

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *314*(1165), 1–340. https://doi.org/10.1098/rstb.1986.0056

Wickersham, I. R., Lyon, D. C., Barnard, R. J. O., Mori, T., Finke, S., Conzelmann, K.-K., Young, J. A. T., & Callaway, E. M. (2007). Monosynaptic Restriction of Transsynaptic Tracing from Single, Genetically Targeted Neurons. *Neuron*, *53*(5), 639–647. https://doi.org/10.1016/j.neuron.2007.01.033

Wickersham, I. R., & Sullivan, H. A. (2015). Rabies Viral Vectors for Monosynaptic Tracing and Targeted Transgene Expression in Neurons [Publisher: Cold Spring Harbor Laboratory Press]. *Cold Spring Harbor Protocols*, *2015*(4), pdb.prot072389. https://doi.org/10.1101/pdb.prot072389

Wickersham, I. R., Sullivan, H. A., & Seung, H. S. (2010). Production of glycoprotein-deleted rabies viruses for monosynaptic tracing and high-level gene expression in neurons. *Nature Protocols*, *5*(3), 595–606. https://doi.org/10.1038/nprot.2009.248

Winding, M., Pedigo, B. D., Barnes, C. L., Patsolic, H. G., Park, Y., Kazimiers, T., Fushiki, A., Andrade, I. V., Khandelwal, A., Valdes-Aleman, J., Li, F., Randel, N., Barsotti, E., Correia, A., Fetter, R. D., Hartenstein, V., Priebe, C. E., Vogelstein, J. T., Cardona, A., & Zlatic, M. (2023). The connectome of an insect brain [Publisher: American Association for the Advancement of Science]. *Science*, *379*(6636), eadd9330. https://doi.org/10.1126/science.add9330

Winnubst, J., Bas, E., Ferreira, T. A., Wu, Z., Economo, M. N., Edson, P., Arthur, B. J., Bruns, C., Rokicki, K., Schauder, D., Olbris, D. J., Murphy, S. D., Ackerman, D. G., Arshadi, C., Baldwin, P., Blake, R., Elsayed, A., Hasan, M., Ramirez, D., … Chandrashekar, J. (2019). Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain. *Cell*, *179*(1), 268–281.e13. https://doi.org/10.1016/j.cell.2019.07.042

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biology*, *19*(1), 15. https://doi.org/10.1186/s13059-017-1382-0

Wu, Z., Miller, E., Agbandje-McKenna, M., & Samulski, R. J. (2006). A2,3 and 2,6 N-Linked Sialic Acids Facilitate Efficient Binding and Transduction by Adeno-Associated Virus Types 1 and 6 [Publisher: American Society for Microbiology]. *Journal of Virology*, *80*(18), 9093–9103. https://doi.org/10.1128/jvi.00895-06

Xu, P., Peng, J., Yuan, T., Chen, Z., He, H., Wu, Z., Li, T., Li, X., Wang, L., Gao, L., Yan, J., Wei, W., Li, C. T., Luo, Z.-G., & Chen, Y. (2024). High-throughput mapping of single-neuron projection and molecular features by retrograde barcoded labeling (J. J. Day & K. M. Wassum, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *13*, e85419. https://doi.org/10.7554/eLife.85419

Yabashi, M., & Tanaka, H. (2017). The next ten years of X-ray science [Publisher: Nature Publishing Group]. *Nature Photonics*, *11*(1), 12–14. https://doi.org/10.1038/nphoton.2016.251

Yang, H. H., Brezovec, B. E., Serratosa Capdevila, L., Vanderbeck, Q. X., Adachi, A., Mann, R. S., & Wilson, R. I. (2024). Fine-grained descending control of steering in walking Drosophila. *Cell*, *187*(22), 6290–6308.e27. https://doi.org/10.1016/j.cell.2024.08.033

Yang, L., Liu, F., Hahm, H., Okuda, T., Li, X., Zhang, Y., Kalyanaraman, V., Heitmeier, M. R., & Samineni, V. K. (2024, April). Projection-TAGs enable multiplex projection tracing and multimodal profiling of projection neurons [Pages: 2024.04.24.590975 Section: New Results]. https://doi.org/10.1101/2024.04.24.590975

Yao, S., Wang, Q., Hirokawa, K. E., Ouellette, B., Ahmed, R., Bomben, J., Brouner, K., Casal, L., Caldejon, S., Cho, A., Dotson, N. I., Daigle, T. L., Egdorf, T., Enstrom, R., Gary, A., Gelfand, E., Gorham, M., Griffin, F., Gu, H., … Zeng, H. (2023). A whole-brain monosynaptic input connectome to neuron classes in mouse visual cortex [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *26*(2), 350–364. https://doi.org/10.1038/s41593-022-01219-x

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

Ye, Q., Gast, G., Su, X., Saito, T., Saido, T. C., Holmes, T. C., & Xu, X. (2022). Hippocampal neural circuit connectivity alterations in an Alzheimer's disease mouse model revealed by monosynaptic rabies virus tracing. *Neurobiology of Disease*, *172*, 105820. https://doi.org/10.1016/j.nbd.2022.105820

Yin, W., Brittain, D., Borseth, J., Scott, M. E., Williams, D., Perkins, J., Own, C. S., Murfitt, M., Torres, R. M., Kapner, D., Mahalingam, G., Bleckert, A., Castelli, D., Reid, D., Lee, W.-C. A., Graham, B. J., Takeno, M., Bumbarger, D. J., Farrell, C., … da Costa, N. M. (2020). A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy [Publisher: Nature Publishing Group]. *Nature Communications*, *11*(1), 4949. https://doi.org/10.1038/s41467-020-18659-3

Young, J. A., Bates, P., & Varmus, H. E. (1993). Isolation of a chicken gene that confers susceptibility to infection by subgroup A avian leukosis and sarcoma viruses [Publisher: American Society for Microbiology]. *Journal of Virology*, *67*(4), 1811–1816. https://doi.org/10.1128/jvi.67.4.1811-1816.1993

Yuan, L., Chen, X., Zhan, H., Henry, G. L., & Zador, A. M. (2024). Massive multiplexing of spatially resolved single neuron projections with axonal BARseq [Publisher: Nature Publishing Group]. *Nature Communications*, *15*(1), 8371. https://doi.org/10.1038/s41467-024-52756-x

Zador, A. M., Dubnau, J., Oyibo, H. K., Zhan, H., Cao, G., & Peikon, I. D. (2012). Sequencing the Connectome. *PLoS Biology*, *10*(10), e1001411. https://doi.org/10.1371/journal.pbio.1001411

Zampieri, N., Jessell, T. M., & Murray, A. J. (2014). Mapping sensory circuits by anterograde transsynaptic transfer of recombinant rabies virus. *Neuron*, *81*(4), 766–778. https://doi.org/10.1016/j.neuron.2013.12.033

Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., Rolny, C., Castelo-Branco, G., Hjerling-Leffler, J., & Linnarsson, S. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, *347*(6226), 1138–1142. https://doi.org/10.1126/science.aaa1934

Zeisler, Z. R., London, L., Janssen, W. G., Fredericks, J. M., Elorette, C., Fujimoto, A., Zhan, H., Russ, B. E., Clem, R. L., Hof, P. R., Stoll, F. M., & Rudebeck, P. H. (2023). High-throughput sequencing of macaque basolateral amygdala projections reveals dissociable connectional motifs with frontal cortex. *bioRxiv: The Preprint Server for Biology*, 2023.01.18.524407. https://doi.org/10.1101/2023.01.18.524407

Zhang, A., Jin, L., Yao, S., Matsuyama, M., Velthoven, C. v., Sullivan, H., Sun, N., Kellis, M., Tasic, B., Wickersham, I. R., & Chen, X. (2024). Rabies virus-based barcoded neuroanatomy resolved by single-cell RNA and in situ sequencing [Publisher: eLife Sciences Publications Limited]. *eLife*, *12*. https://doi.org/10.7554/eLife.87866.2

Zhang, M., Pan, X., Jung, W., Halpern, A. R., Eichhorn, S. W., Lei, Z., Cohen, L., Smith, K. A., Tasic, B., Yao, Z., Zeng, H., & Zhuang, X. (2023). Molecularly defined and spatially resolved cell atlas of the whole mouse brain [Publisher: Nature Publishing Group]. *Nature*, *624*(7991), 343–354. https://doi.org/10.1038/s41586-023-06808-9

Zheng, M., Hu, Z., Mei, X., Ouyang, L., Song, Y., Zhou, W., Kong, Y., Wu, R., Rao, S., Long, H., Shi, W., Jing, H., Lu, S., Wu, H., Jia, S., Lu, Q., & Zhao, M. (2022). Single-cell sequencing shows cellular heterogeneity of cutaneous lesions in lupus erythematosus [Publisher: Nature Publishing Group]. *Nature Communications*, *13*(1), 7489. https://doi.org/10.1038/s41467-022-35209-1

Zuber, B., Nikonenko, I., Klauser, P., Muller, D., & Dubochet, J. (2005). The mammalian central nervous synaptic cleft contains a high density of periodically organized complexes [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *102*(52), 19192–19197. https://doi.org/10.1073/pnas.0509527102