

Imperial College London

Department of Life Sciences

**INTERROGATING NEURONAL CONNECTIVITY THROUGH SERIAL TWO-PHOTON  
TOMOGRAPHY OF THE WHOLE MOUSE BRAIN EX-VIVO**

*Author:*

Gerald Moore

*Supervisors:*

Prof. Stephen Brickley

Prof. Simon Schultz

A thesis submitted for the degree of Doctor of Philosophy

Imperial College London

April, 2021

## ABSTRACT

Ex-vivo, serial two-photon (STP) tomography was used to image the entire mouse brain to interrogate neuronal connectivity across local and distal brain volumes. Large terabyte-sized data sets presented many analytical challenges. To this end, the following pipelines were created and validated; stitching of tile image data, registration of data with common coordinate framework, and a deep learning cell detection algorithm that was comparable to human performance. These pipelines were developed to facilitate two scientific studies. First, the unknown connectivity of local interneurons of the dorsal lateral geniculate nucleus (LGd) of Sox14cre/+ mice were mapped using rabies virus tracing, revealing previously unreported input from the GABAergic reticular nucleus (RT), suggesting a double inhibition mechanism of modulation onto local thalamocortical relay cells. In-vitro patch-clamp recordings confirmed functional connectivity. Thalamic interneurons in the Sox14gfp/+ mouse had high prevalence in first order (FO) sensory nuclei, challenging the notion that interneurons in the mouse thalamus are constrained to visual thalamus only. In-vitro patch-clamp recordings showed functional interneuron-interneuron connectivity as suggested by interneuron morphological analysis. Second, STP tomography and magnetic resonance imaging (MRI) were explored for correlating age-related cognitive changes with anatomical connectivity. Female C57BL/6J mice were subjected to a novel nine-arm radial maze task, and object location and recognition tasks at 4, 8, 12 and 18 months of age. Mice showed remarkable retention of cognitive performance throughout age (n=58). Unique arm choice presented one of many measures that showed performance predictability at 18 months given 4 months performance ( $R^2=0.6166$ ,  $p<0.0001$ ,  $n=22$ ). A combination of unique arm sequence length, cumulative arm time and velocity measures showed improved predictability for 18 months performance given 4 months performance ( $R^2=0.6239$ ,  $p<0.0001$ ,  $n=22$ ). This level of predictability has been previously seen in Lothian cohort human studies. STP tomography and MRI were used to demonstrate mapping brain connectivity changes with age.



## TABLE OF CONTENTS

|                                                                                                      |     |
|------------------------------------------------------------------------------------------------------|-----|
| List of Figures and Tables                                                                           | 5   |
| List of Abbreviations                                                                                | 8   |
| Declaration of Originality                                                                           | 11  |
| Copyright Declaration                                                                                | 12  |
| Acknowledgements                                                                                     | 13  |
| Chapter 1 – Introduction                                                                             | 15  |
| 1.1 Foreword                                                                                         | 16  |
| 1.2 Background                                                                                       | 16  |
| 1.3 High-resolution volume-wide imaging                                                              | 18  |
| 1.4 Serial imaging                                                                                   | 21  |
| 1.5 Challenges in understanding whole brain data                                                     | 23  |
| 1.6 Staining and viral assays                                                                        | 24  |
| 1.7 Research focus                                                                                   | 26  |
| Chapter 2 – High-resolution serial two-photon tomography for ex-vivo whole brain imaging             | 31  |
| 2.1 Introduction                                                                                     | 32  |
| 2.2 Methods                                                                                          | 39  |
| 2.3 Results                                                                                          | 46  |
| 2.4 Discussion                                                                                       | 54  |
| 2.5 Troubleshooting, caveats and future perspectives                                                 | 56  |
| Chapter 3 – Whole-brain mouse registration to common coordinate framework for anatomical delineation | 60  |
| 3.1 Introduction                                                                                     | 61  |
| 3.2 Methods                                                                                          | 75  |
| 3.3 Results                                                                                          | 83  |
| 3.4 Discussion                                                                                       | 94  |
| 3.5 Future perspectives                                                                              | 96  |
| 3.6 Supplementary data                                                                               | 97  |
| Chapter 4 – Deep learning cell counting in anatomically targeted structures                          | 102 |

|                                                                                                                                                                     |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.1 Introduction                                                                                                                                                    | 103 |
| 4.2 Methods                                                                                                                                                         | 113 |
| 4.3 Results                                                                                                                                                         | 125 |
| 4.4 Discussion                                                                                                                                                      | 139 |
| 4.5 Future perspectives                                                                                                                                             | 142 |
| 4.6 Supplementary data                                                                                                                                              | 144 |
| Chapter 5 – Interneurons widely populate many nuclei in the mouse thalamus and are connected to the reticular nucleus and other interneurons of the visual thalamus | 145 |
| 5.1 Introduction                                                                                                                                                    | 147 |
| 5.2 Methods                                                                                                                                                         | 153 |
| 5.3 Results                                                                                                                                                         | 161 |
| 5.4 Discussion                                                                                                                                                      | 182 |
| 5.5 Future perspectives                                                                                                                                             | 186 |
| 5.6 Supplementary data                                                                                                                                              | 188 |
| Chapter 6 – Age-related changes in cognitive performance and brain anatomy in C57BL/6J mice from young to old age                                                   | 189 |
| 6.1 Introduction                                                                                                                                                    | 190 |
| 6.2 Methods                                                                                                                                                         | 196 |
| 6.3 Results                                                                                                                                                         | 204 |
| 6.4 Discussion                                                                                                                                                      | 222 |
| 6.5 Future directions                                                                                                                                               | 226 |
| 6.6 Supplementary data                                                                                                                                              | 228 |
| Bibliography                                                                                                                                                        | 230 |

## LIST OF FIGURES AND TABLES

|                                                                                                                                                      |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 2.1.</b> TissueCyte 1000, for serial-two photon imaging across tissue volumes                                                              | 39  |
| <b>Figure 2.2.</b> Image stitching pipeline                                                                                                          | 47  |
| <b>Figure 2.3.</b> Exemplar results for serial two-photon whole brain imaging and stitching                                                          | 50  |
| <b>Figure 2.4.</b> pAAV-CAG-tdTomato (anterograde) and pAAV-hSyn-EGFP (retrograde) co-injection labelling of LGd connectivity                        | 52  |
| <b>Figure 3.1.</b> Human manual segmentation error for nine test anatomical regions in the mouse brain                                               | 64  |
| <b>Figure 3.2.</b> Principles of registration                                                                                                        | 67  |
| <b>Figure 3.3.</b> Registration pipeline of STP data to Allen CCF3 and segmentation using the annotation atlas                                       | 84  |
| <b>Figure 3.4.</b> Transformation of sparse points for target registration error (TRE) evaluation in the four tested parameter map configurations    | 86  |
| <b>Figure 3.5.</b> Human target error assessment                                                                                                     | 87  |
| <b>Figure 3.6.</b> Qualitative registration result                                                                                                   | 90  |
| <b>Figure 3.7.</b> Similarity measures by correlation coefficient (CC) and mutual information (MI)                                                   | 93  |
| <b>Supplementary table 1.</b> Rigid transformation parameter configurations for Par00GM, Par0025, Par0033 and Par0038                                | 98  |
| <b>Supplementary table 2.</b> Affine transformation parameter configurations for Par00GM, Par0025, Par0033 and Par0038                               | 99  |
| <b>Supplementary table 3.</b> B-spline transformation parameter configurations for Par00GM, Par0025, Par0033 and Par0038                             | 100 |
| <b>Figure 4.1.</b> U-Net architecture                                                                                                                | 108 |
| <b>Figure 4.2.</b> Curation of training data for the U-Net                                                                                           | 127 |
| <b>Figure 4.3.</b> Evaluating loss functions using faux simulated data sets with eight predictions of increasing intersection-over-union (IoU) score | 129 |
| <b>Figure 4.4.</b> Hyperparameter tuning                                                                                                             | 132 |
| <b>Figure 4.5.</b> Trained U-Net evaluation using novel images                                                                                       | 134 |

|                                                                                                                                                                                |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 4.6.</b> Complete counting pipeline using registration to Allen CCF3 atlases to target automated cell counting to anatomical regions of interest                     | 136 |
| <b>Supplementary Figure 4.1.</b> Performance times for a variety of Python image processing libraries for image reading, gaussian filter operation and median filter operation | 144 |
| <b>Figure 5.1.</b> STP imaging across the thalamus of a Sox14gfp/+ mouse                                                                                                       | 162 |
| <b>Figure 5.2.</b> Qualitative comparison of Sox14gfp/+ and Sox14gfp/gfp mice at AP - 4.1                                                                                      | 165 |
| <b>Figure 5.3.</b> Quantitative mapping of GFP+ neurons in Sox14gfp/+ and Sox14gfp/gfp mice                                                                                    | 166 |
| <b>Figure 5.4.</b> Presynaptic mapping to Sox14 expressing cells in the LGd                                                                                                    | 169 |
| <b>Figure 5.5.</b> Mapping visual cortex inputs to LG                                                                                                                          | 171 |
| <b>Figure 5.6.</b> Whole-cell voltage-clamp recordings of Sox14 interneurons of the LGd following stimulation of RT                                                            | 172 |
| <b>Figure 5.7.</b> Segmenting paired interneuron biocytin fills using deep learning                                                                                            | 176 |
| <b>Figure 5.8.</b> Detecting putative interneuron-interneuron anatomical connectivity                                                                                          | 177 |
| <b>Figure 5.9.</b> Paired interneuron recording in the LGd                                                                                                                     | 181 |
| <b>Supplementary Figure 5.1.</b> Widefield imaging of Alexa 555 conjugated biocytin fills of interneurons                                                                      | 188 |
| <b>Figure 6.1.</b> Nine-arm radial maze                                                                                                                                        | 204 |
| <b>Figure 6.2.</b> Body weight and open field testing                                                                                                                          | 206 |
| <b>Figure 6.3.</b> Radial maze and object recognition and location performance                                                                                                 | 208 |
| <b>Figure 6.4.</b> Tracking individual animal performance with age                                                                                                             | 215 |
| <b>Figure 6.5.</b> Combining performance metrics for individual animal performance using standardised Z-score to reflect individual animal performance                         | 218 |
| <b>Figure 6.6.</b> Grey matter and white matter segmentation from MRI T1 scans (left) using FSL                                                                                | 219 |
| <b>Figure 6.7.</b> Demonstration of high-resolution connectivity mapping in the PFC following pAAV-hSyn-EGFP (retrograde) virus injection                                      | 220 |
| <b>Supplementary Figure 6.1.</b> Skewness measurements for each radial maze and object recognition and location task for the population                                        | 228 |

**Supplementary Figure 6.2.** Coefficient of variation (CV) measurements for each radial maze and object recognition and location task for the population 229

## LIST OF ABBREVIATIONS

The following is a list of nomenclature that covers the complete thesis. Of note, abbreviations for anatomical nuclei follow the nomenclature of the Allen Brain Institute but are frequently interchanged between published work. For example, the dorsal path of the lateral geniculate nucleus is abbreviated as LGd in the Allen Brain Institute nomenclature and thus in this thesis but is most commonly written as dLGN in published work. *Id est* (i.e.) variants are added to the definition below for clarification and familiarity across different literature.

2P; two-photon (i.e. TP)

3P; three-photon

AAV; adeno-associated virus

ACA; anterior cingulate area

AD; Alzheimer's Disease

AHN; anterior hypothalamic nucleus

AMMI; advanced Mattes mutual information

ANC; advanced normalised correlation

ANN; artificial neural network

AOD; acousto-optical device

A $\beta$ ; amyloid-beta

BCE; binary cross-entropy

CC; correlation coefficient

CCF3; Allen Mouse Brain Common Coordinate Framework atlas (version 3)

CLSM; confocal light sheet microscopy

CNN; convolution neural network

CSF; cerebrospinal fluid

CSF: cerebrospinal fluid

CV; coefficient of variability

dACA; dorsal anterior cingulate area

DL; deep learning

DTI; diffusion tensor imaging

EM; electron microscopy

EnvA; avian sarcoma leucosis virus glycoprotein  
fMRI; functional magnetic resonance imaging  
FOV; field-of-view  
FP; Franklin and Paxinos mouse brain atlas  
FSL; FMRIB software library  
FTL; focal Tversky  
G; glycoprotein  
GMM; Gaussian mixture model  
HAROLD; hemispheric asymmetry reduction in older adult  
HIV; human immunodeficiency virus  
HPF; hippocampus formation  
HSV-1; herpes simplex virus type 1  
IGL; intergeniculate leaflet  
IL; infralimbic  
IoU; intersection over union (i.e. Jaccard)  
IP; intraperitoneal  
IPSC; inhibitory postsynaptic current (mini, mIPSC; spontaneous, sIPSC)  
IT; intratelencephalic  
LG; lateral geniculate nucleus (i.e. LGN)  
LGd; dorsal part of the lateral geniculate nucleus (i.e. dLGN)  
LGv; ventral part of the lateral geniculate nucleus (i.e. vLGN)  
LSM; light sheet microscopy  
MI; mutual information  
MOST; micro-optical sectioning tomography  
MRI; magnetic resonance imaging  
MV; medial vestibular nucleus  
NA; numerical aperture  
NAS; network attached storage  
PBS; phosphate buffered saline  
PFA; paraformaldehyde  
PFC; prefrontal cortex  
PL; prelimbic

PMT; photo-multiplier tube  
PRV; pseudo-rabies virus  
PSTH; peristimulus time histogram  
RGC; retinal ganglion cell  
RSP; retrosplenial cortex  
RT; reticular nucleus (i.e. RTN)  
RV; rabies virus  
RVdG; glycoprotein-deleted rabies virus  
SC; superior colliculus  
SEM; scanning electron microscopy  
SIFT; scale-invariant feature transform  
SNR; signal-to-noise ratio  
SSp; primary somatosensory area  
SSP; somatosensory cortex  
STP; serial two-photon  
SUB; subiculum  
TC; thalamocortical  
TMS; transcranial magnetic stimulation  
TRE; target registration error  
TVA; cellular receptor for avian sarcoma leucosis virus glycoprotein  
vACA; ventral anterior cingulate area  
VIS; visual cortex  
VISam; anteromedial part of the secondary visual cortex  
VISp; primary visual cortex  
VISp; primary visual cortex (i.e. V1)  
VPM; ventral posteromedial nucleus of the thalamus  
WMH; white matter hyperintensities



## **DECLARATION OF ORIGINALITY**

I, Gerald Moore, hereby certify that I am the sole author of this thesis and that this work has not been previously submitted for any other qualification. Any sources of published or unpublished work that has been referenced in this thesis has been appropriately acknowledged either in the bibliography or directly in text. Any contributing effort from colleagues has been appropriately acknowledged in the relevant methods sections.

## **COPYRIGHT DECLARATION**

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to supervisors Prof. Stephen Brickley of the Department of Life Sciences and Prof. Simon Schultz of the Department of Bioengineering for giving me the opportunity to undertake this research and providing me the means and encouragement to do so. Both supervisors have excelled in cultivating extremely supportive and friendly research groups. Being engrained in both groups really elevated the notion that research is a collaborative effort. Prof. Stephen Brickley was inexhaustible as a primary supervisor and offered extensive patience and enthusiasm at every step of the research. Whilst many do not get the opportunity or would be dissuaded from the fact, having a desk positioned right outside his office door was an absolute gift. I will not forget the number of work-related or off-hand discussions covering any number of exotic topics.

I would also like to thank all the members Brickley lab and Schultz lab. In particular I am grateful for Rushdie Abuhamdah for his endless wisdom and dedication to his work and for providing the one person in the entire office who cared about sports, albeit the wrong teams. I am also thankful for the brief presence of Florian Zirpel who provided many interesting discussions, medical or otherwise. Also, Jeevan Soor, for his unyielding happiness, unparalleled kind-heartedness and consideration to all those around him. Lastly, Diana Lucaci for providing years of wonderfully frank discussions and for thinking apple tasting was a good activity to do with undergraduate students.

Of course, I almost certainly wouldn't be here without the encouragement and love of my parents. Even if my mother still does not understand what I do, my father's scientific career and fascination is almost certainly responsible for igniting my own. Likewise, I would like to thank my brother, Robert, for providing an element of brotherly joy in my life. Crucially however, for having three majestic cats for me to visit, Groot, Obi and Indy, for which the last cat proved valuable as subject for my image registration figure (Chapter 3, Fig. 3.1).

The last person I would like to thank is my partner, Aimee Laura, for providing unbridled delight in my life. She is solely responsible for keeping me sane during my research, during writing, and during a time where the world succumbed to COVID-19. I am forever grateful

that she understood and helped ease the stresses and strains that come with scientific research and life as a whole.

Lastly, I would like to thank the existence of bouldering and rock climbing for providing me a much-needed outlet for stress on an almost-daily basis. I would not like to thank COVID-19 for taking that away from me in recent months...

## Chapter 1

### **INTRODUCTION**

## **1.1 FOREWORD**

The work presented in this thesis covers a range of topics, including engineering aspects with novel microscopy imaging techniques and development, computational aspects with image registration, segmentation and machine-learning-based challenges, and biological aspects including mapping structural and functional connectivity in the mouse brain and behavioural experimentation for assessment of age-related cognitive decline. This chapter serves as a general-purpose introduction to cover important background information, with each chapter beginning with a more specific introduction. Fundamentally, the goal of this thesis is to address questions in two areas. The first area concerns the visual circuitry system and the connectivity and function of local interneurons found in the thalamic region of the brain. The second area concerns cognitive progression in healthy ageing mice and the associated anatomical correlates. Here, we lay the foundation of the background knowledge, techniques and research approach for addressing these topics in the brain of a small mammalian rodent, the humble mouse.

## **1.2 BACKGROUND**

One of the unrelenting challenges throughout the history of neurobiology, and the general field of biology, is understanding the interconnection between structure and function of the mammalian brain. Neuroanatomy's godfather, Santiago Ramón y Cajal, pioneered our understanding of the central nervous system by deconstructing the perceived notion of a single continuous wiring system into individual functional units (Ramón y Cajal, 1909; Ramón Y Cajal, 1894). These units later became known as 'Neurons'. Research from Luigi Galvani and Volta in the latter half of the 18<sup>th</sup> Century demonstrated that the conduction and transmission of electrical impulses enabled information propagation through an interconnected network of neurons in the central nervous system (Bresadola, 1998). Hodgkin and Huxley fundamentally expanded on this mechanism by constructing a mathematical model that governed the electrical signal generation and propagation (i.e. action potentials) through ion movement, by using squid giant axons in 1952 (Hodgkin & Huxley, 1952).

Since then, there has been extensive research and breakthroughs in the field of neurobiology, covering aspects of brain structure and also function. For example; the development of the patch-clamp technique for measuring ionic currents in single cells and single ion channels

(Neher & Sakmann, 1992); the formulation of the split-brain theory and idea that, animals with unconnected or severed hemispheres exhibit cerebral localisation of different cognitive and language functions (Pearce, 2019); and recent work demonstrating that paralysed chronic motor complete spinal cord injury patients can regain walking function following spinal cord epidural stimulation (Angeli et al., 2018). Nevertheless, our understanding of the brain is still far from complete. The brain and its interlinked components are staggeringly abundant and diverse. The human brain is composed of 170 billion neuronal and non-neuronal cells, despite weighing only 1.3 kg and amounting a volume of 1,200 cm<sup>3</sup> (Azevedo et al., 2009). For context, there exists an estimated 100 billion stars in the Milky Way, and the most recent attempt to chart the positions of these stars have focused on just one-per-cent of these. This charting survey started in 2014 and is only due to finish at the end of 2020 (Prusti et al., 2016); a six-year endeavour.

Even tiny organisms show immense complexity. The nematode roundworm (*Caenorhabditis elegans*; *C. elegans*) at only 1 mm in length, has a nervous system composed of ~300 almost completely structurally and functionally unique neurons, differentiable into as many as 118 classes based on morphology and connectivity. There can be as many as 8000 synapses in total (White et al., 1986). Smaller organisms such the *C. elegans*, fruit fly (*Drosophila*), zebrafish (*Danio rerio*) and the rodent mouse, are valuable for experimental research due to reduced complexity, smaller size for more convenient housing and breeding, and more straightforward genetic mapping and manipulation. Indeed, the entire genome sequence of the *C. elegans* was published back 1998 (The *C. elegans* Sequencing Consortium, 1998), the zebrafish in 2013 (Howe et al., 2013) and *Drosophila* in 2000 (Adams et al., 2000).

Although replication of human disease in animal models can be challenging and limited, they provide an essential tool for studying the underlying mechanisms and genetics of diseases. *C. elegans* present amyloid precursor protein orthologues, making them valuable for studying amyloid-beta (A $\beta$ ) toxicity in Alzheimer's disease (AD) research (Calahorra & Ruiz-Rubio, 2011). Additionally, *C. elegans* mutants expressing human genes related to Parkinson's disease allows studies on the degeneration of dopaminergic neurons (Calahorra & Ruiz-Rubio, 2011), behavioural aspects of the disease, and environmental influences (A. J. Harrington et al., 2010). The phenotypic transparency of zebrafish allows them to be imaged at high-resolution,

*in-vivo*, and over an extended period. They, therefore, provide a convenient model for studying the development of the cardiovascular system, angiogenesis which has vital implications in tumour growth (Chávez et al., 2016), and regenerative biology (Azevedo et al., 2011). The *Drosophila* genome contains homologs for 75% of genes responsible for human disease, making them valuable for cancer research (Mirzoyan et al., 2019). Lastly, despite the larger size, mice constitute 73% of animal usage in experimental and breeding procedures in 2018 (Home Office, 2019). Increased neuronal wiring complexity makes them ideal for studying complex behaviours and physiology. At the turn of the millennia, the Nobel prize-winning research of Mario R. Capecchi, Martin J. Evans and Oliver Smithies for genetic modification in mice, resulted in rapid interest in gene targeting studies. Knock-out, knock-in and transgenic mouse models have proven indispensable for extensive areas of biomedical research. For example, transgenic mice deficient in the p53 gene showed the importance of the gene for the predisposition of tumour growth (Donehower et al., 1992), and were also crucial for revealing the dissemination of cancer cells in early stages of mammary carcinoma (Hüsemann et al., 2008). Research into neurodegenerative diseases, such as Huntington's (Cepeda-Prado et al., 2012) and amyotrophic lateral sclerosis (De Giorgio et al., 2019), have also been greatly furthered by genetic manipulation of the mouse.

A valuable property of these organisms is that they are tiny. Their size allows for more convenient imaging at high-resolution, using techniques such as light-based microscopy and electron microscopy (EM). Given volume of the human brain, these techniques would otherwise be extraordinarily difficult to utilise without greatly restricting the volume of interest. Indeed, *C. elegans*, the larval brain of zebrafish and the brain of *Drosophila* have been imaged at nanometre resolution using serial scanning electron microscopy (SEM; Hildebrand et al., 2017; Müller-Reichert et al., 2008; Zheng et al., 2017). However, serial SEM imaging can still take many days or weeks to accomplish in small tissue volumes (Hall et al., 2012). The complete EM scan of the *Drosophila* brain at 0.08 mm<sup>3</sup> voxel resolution took three weeks to accomplish and took longer to process and analyse (Zheng et al., 2017). EM reconstruction of the whole mouse brain is therefore entirely unfeasible at this current point in time. Nevertheless, fast high-resolution imaging of large volumes of tissue provides the cornerstone for anatomical connectivity research and is therefore important in the context of this thesis.



### 1.3 HIGH-RESOLUTION VOLUME-WIDE IMAGING

Imaging in neurobiology faces three fundamental challenges. First, neurons are microscopic with soma between 10 and 200  $\mu\text{m}$  in diameter, dendrites as small as 1  $\mu\text{m}$ , and dendritic spines  $<1$   $\mu\text{m}$  (Kawaguchi & Kubota, 1998). Imaging techniques must, therefore, provide sufficient resolution to distinguish between structures on the submicron scale. Second, neurons in the brain scatter across all three dimensions, span local and distal regions, and can have arbourisations which accumulate to many thousands of microns in length (Morgan & Lichtman, 2020; S. Reardon, 2017). Imaging systems must, therefore, offer the capability to image both wide-and-deep into the tissue. Third, of particular importance to *in-vivo* physiological studies, transient membrane potentials and currents occur over millisecond timescales, necessitating imaging techniques with a high temporal-resolution to capture these events. Light-based fluorescence imaging modalities including wide-field microscopy, confocal microscopy, two-photon (2P) microscopy (and furtherance multi-photon techniques) and light-sheet microscopy establish an impressive array of instruments for cellular and subcellular imaging. Despite light-based imaging techniques addressing these fundamental challenges, no imaging techniques exist that excel at addressing all three.

Exploiting and circumventing characteristics of intrinsic or genetically manipulated expression of fluorophores in tissue allows for increased resolution imaging. When combined with a high numerical aperture (NA) objective, wide-field microscopy offers fast, high-resolution and sensitive imaging (Kner et al., 2010). However, high sensitivity in a wide-field configuration is only possible as all emitted photons are detected. Imaging deeper than 100  $\mu\text{m}$  underneath the tissue surface, results in the high detection of out-of-focal-plane photons. Refractive index mismatch within the sample, and between the sample itself and the objective, further restrict wide-field imaging to superficial depths.

Confocal microscopy, conceptualised in the 1950s, revolutionised fluorescence microscopy by preventing photons originating from outside the focal plane from reaching the detector (Minsky, 1988). The design, therefore, improved resolution in the spatial and axial dimensions and also introduced optical sectioning, a technique which allows sampling of an axially confined plane deep into tissue without collecting photons from anywhere else (Jonkman & Brown, 2015). The benefits of the confocal configuration were refined through multi-photon

approaches. 2P microscopy achieves confocality through the simultaneous absorption of two photons with each photon quantised with half the energy required for fluorescence excitation (Denk et al., 1990; Helmchen & Denk, 2005), a phenomenon predicted back in 1930 (Göppert-Mayer, 2009). As the cross-section for photon-photon interaction has a quadratic relationship with the density of photons, the excitation beam is axially confined to a focal volume of around  $0.1 \mu\text{m}^3$  (Zipfel et al., 2003). A smaller excitation volume has the advantage of reducing photobleaching and phototoxicity compared to a confocal system which illuminates an axial volume above and below the focal point. A lower energy laser also means reduced Rayleigh scattering allowing optical sectioning at deeper depths into biological tissue (So et al., 1998). This is particularly valuable for *in-vivo* applications where imaging as deep as 1.6 mm in the mouse cortex is achievable in intact tissue (Kobat et al., 2011). However, confocal and 2P methods are limited in temporal-resolution as they require mechanically point-scanning the focal volume across the specimen. The spinning disc design generates an array of focal volumes that are simultaneously scanned across the field-of-view and offers a faster alternative to single-point laser scanning configuration (Petráň et al., 1968). The design was first explored with the confocal microscope (Schulz et al., 2013) and has been adapted to 2P microscopes (Maeda et al., 2018; Otomo et al., 2015).

Three-photon (3P) microscopy offers improved imaging depths over 2P and confocal systems with studies of the hippocampus in intact brains (Horton et al., 2013; Ouzounov et al., 2017; T. Wang et al., 2018) and even demonstration of imaging through the intact mouse skull (T. Wang et al., 2018). However, exceptionally powerful lasers are required to compress enough photons into the focal volume to trigger three-photon absorption events.

Post-imaging deconvolution techniques can also improve instrument resolving power by removing noise and blurring through modelling photon detection using the point spread function of the system (Sibarita, 2005). Such approaches can improve the image signal-to-noise ratio (SNR) without necessitating drastic modifications to existing set-ups. They can also improve systems which already use confocality (Noakes et al., 2001).

Ultimately, the resolving power of light-based imaging methods is constrained by the diffraction limit, which is in itself related to wavelength. As EM operates with wavelengths

$10^5$  times smaller than visible light, EM systems offer a powerful technique for sub-nanometre resolution imaging (Leonard et al., 2012). Despite having a much greater depth-of-field than an optical microscope, EM microscopy offers no ability for sub-surface imaging in intact tissue.

In some cases, speed might be the essential aspect of an imaging system. In an *ex-vivo* experiment, imaging large volumes of tissue could be time-intensive. A confocal microscope, for example, can take half a day to image a  $0.16 \text{ mm}^3$  volume of tissue (see 'Chapter 4'). Although automated systems can typically be left alone, extended imaging times can lead to issues with specimen drift and evaporation or loss of immersion media for high NA objectives.

For *in-vivo* experiments using temporal probes such as calcium transient imaging with G-CaMP (Nakai et al., 2001), high temporal-resolution with millisecond precision and depth imaging is required to capture neuronal activity across small volumes of tissue (Russell, 2011). One strategy to address low imaging speed is through acousto-optical devices (AODs) which permit rapid refocusing of the laser between pre-determined positions in the field-of-view; maximising signal integration on structures of interest (Grewe et al., 2010). Such systems can accomplish 1 kHz sampling rates. Single AODs are typically limited to single focal plane operation. However, several AODs can be combined for three-dimensional inertia-free scanning (Reddy et al., 2008). Temporal focusing can also provide faster acquisition rates with 2P resolution imaging, by removing the necessity to scan a single focal point (Isobe et al., 2017; Oron et al., 2005; Papagiakoumou et al., 2020). Temporal scanning in the axial axis can be used for depth-resolved imaging (Oron et al., 2005) and has been demonstrated for fast *in-vivo* calcium imaging (Prevedel et al., 2016).

#### **1.4 SERIAL IMAGING**

The depth characteristics of imaging systems discussed above refer to the inherent depth properties of intact tissue volumes. However, in an *ex-vivo* setting such as those presented in the research here, volumetric imaging can be accomplished using a serial sectioning technique in combination with block-face prepared samples. Serial SEM systems have been used for a complete reconstruction of *C. elegans* at the synaptic level (Hall et al., 2012). However, they can also be used for larger volumes such as *Drosophila* (Zheng et al., 2017), or

targeted to small regions of larger mammal tissue volumes, such the (Morgan et al., 2016; Morgan & Lichtman, 2020). Serial SEM is limited by a long acquisition time, and the amount of generated data in the order of several terabytes. Light-sheet microscopy offers fast optical sectioning by focusing a single sheet of light through a sample combined with wide-field detection (Girkin & Carvalho, 2018; P. J. Keller & Dodt, 2012; Osten & Margrie, 2013). However, samples must be intrinsically transparent or made transparent through tissue clearing techniques such as CLARITY (Chung & Deisseroth, 2013), DISCO, and its variants (Ertürk et al., 2012; Pan et al., 2016; Renier et al., 2014), and CUBIC (Susaki et al., 2014). Unfortunately, clearing can distort and disrupt structures in the tissue (Chung et al., 2013; Richardson & Lichtman, 2015; Silvestri et al., 2016). It can also require specially designed compatible light microscopes and high-NA objectives with a long working distance, and imaging using a refractive-index-matching solution which can be expensive. Additionally, clearing can take several days to achieve with variable success (Silvestri et al., 2016). However, cleared mouse brains can be imaged over a few hours at a resolution that permits the identification of axons and dendrites.

Micro-optical sectioning tomography (MOST; Gong et al., 2013; Li et al., 2010), serial two-photon (STP) tomography (Amato et al., 2016; Ragan et al., 2012) and wide-field large-volume tomography (Gong et al., 2016) can be used to generate large digital data sets of the entire mouse brain at high resolution using physical sectioning and microscopy. Wide-field detection in light-sheet microscopy makes it a fast acquisition method compared to slower laser point-scanning detection methods intrinsic to confocal and 2P techniques (Girkin & Carvalho, 2018). STP tomography is therefore slow for acquisition yet offers high-spatial-resolution down to 0.5  $\mu\text{m}$  which is more than sufficient for imaging axons and dendrites. High axial confinement in optical sectioning combined with an automated microtome for block-face sectioning makes this technique powerful for acquiring high-resolution digital data of large volumes. MOST acquisition is achieved by line-scanning through a bright-field microscope. However, rather than slice complete sections of tissue from the block-face preparation, MOST uses knife-edge sectioning to slice ribbons of tissue alongside simultaneous imaging. Imaging can still take weeks to accomplish for high-resolution studies (Li et al., 2010). Ribbon sectioning is more destructive than whole block-face sectioning approaches meaning slices are unsuitable for post hoc analysis. Of the serial imaging approaches, STP imaging offers an ideal compromise

for high-resolution and reasonable throughput imaging with the capacity to extract slices for post hoc analysis.

Any large volume imaging technique ultimately generates terabyte data sets for mouse brain samples. Although, computing power and availability of high-performance resources such as Amazon Web Services and Google Cloud Computing Services are becoming increasingly accessible, analysis of large imaging data sets is far from trivial. For the scope of this thesis, we were required to address this challenge through a novel perspective for the following considerations.

### **1.5 CHALLENGES IN ANALYSING WHOLE BRAIN DATA**

The establishment of magnetic resonance imaging (MRI) as a flagship technique for extensive volume imaging in human subjects led to a rapid interest in the field of image registration and segmentation. More frequent studies with large data sets meant researchers were required to understand and interrogate the data which might otherwise be limited to expertly trained specialists. Analysis might include recognising and delineating brain areas of interest. However, manual delineation in MRI data is laborious and highly subjective even amongst experts. This issue not only extends to studies which acquire digital data sets of the mouse brain but arguably can be worsened due to high-resolution systems requiring higher precision and accuracy during delineation tasks (Nestor et al., 2013; Niedworok et al., 2016). Mouse brain atlases such as the Franklin and Paxinos atlas (FP; Franklin & Paxinos, 2008), Waxholm Space atlas (Johnson et al., 2010) and various iterations of the Allen Brain Institute atlas including the recently released Common Coordinate Framework atlas (CCF3; Dong, 2008; Q. Wang et al., 2020), provide valuable reference space and common standardised framework. Problematically, not all these atlases agree in terms of anatomical structure labelling and delineation breakdown (Chon et al., 2019). The Allen Brain Institute atlases are also more comprehensive with the demarcation of thousands of anatomical regions and subregions. The Waxholm Space atlas, on the other hand, is limited in region delineation having originated from low-resolution MRI data. The CCF3 atlas has an isotropic ten-micron voxel resolution meaning its sampling in the rostral-caudal extent is higher than other atlases and is more suited for the analysis of complete whole brain data sets. Stereotaxic studies typically use

these atlas frameworks as a reference for delineating regions of interest, although many toolkits now exist automatically register image data with an atlas space.

ANTs (Avants et al., 2014), Elastix (Klein et al., 2009) and SimpleITK (Lowekamp et al., 2013) offer some of the most commonly used registration algorithms. Modern deep-learning segmentation approaches are also gaining more traction and are likely to succeed the intensity-based registration algorithms, as they offer faster, more accurate registration and segmentation results (Iqbal et al., 2019; Tan et al., 2020; Todorov et al., 2020). Problematically, comprehensive approaches for registration of whole mouse brain data sets are lacking, with automated whole mouse brain registration focusing on MRI data (Bink, 2016; Hammelrath et al., 2016; Mulder et al., 2017).

Many studies also seek to quantify structural information such as cell counts in stained or genetically expressed fluorophore labelled tissue. Manual and even automated techniques are subjective and prone to disagreements between independent studies (D. Keller et al., 2018). Manual counting is also challenging and time-intensive. If working with whole mouse brain imaging data, manual counting is near impossible to accomplish. Robust automated cell counting in large microscopy data sets is therefore highly sought after. Some recent developments exist such as ClearMap, designed with cleared tissue in mind (Renier et al., 2016), the brain cell finder tool (Frasconi et al., 2014), and DALMATIAN for use in serial and individually sectioned microscopy data sets (Shuvaev et al., 2017). However, all these techniques suffer from the same limitations; generalisability. Data sets from imaging systems are not standardised between labs, and different imaging systems have different SNR, intensity of intrinsic background fluorescence, morphology of fluorescence signal depending on fluorophore localisation, and resolution. High-resolution whole-brain methods like STP tomography can be used for imaging neurite morphology, but this can easily be mistaken for somatic signal in any image thresholding-based cell counting technique. Deep-learning offers an exciting methodology for learning morphological features about cells that make detection algorithms better suited over intensity-based cell detection methods (Ronneberger et al., 2015; Segebarth et al., 2018). We therefore use the recent developments in deep-learning as part of an end-to-end analysis pipeline to provide a novel analysis strategy for analysing multi-terabyte image data sets.

## 1.6 STAINING AND VIRAL ASSAYS

The development of genetically modified viral tracers and transgenic mouse lines has assisted the interrogation of animal connectomics. These approaches aid visualisation of neuronal morphology and circuits by introducing contrasting agents into structures of interest; a field which rapidly evolved following the revolution of green fluorescent protein (GFP; Chalfie et al., 1994). Initial tracer design required impregnating lesioned axons with silver (Hoff, 1932) with later refinement leading to amino acid based directional tracers (Saleeba et al., 2019). Modern designs now rely on recombinant viral vectors to deliver and drive expression of fluorophores like GFP in transduced neurons. This approach has a major advantage of control over ubiquitous or cell-specific expression based on the types of the promoter used (Y. Wang et al., 2011). Transgenic mouse lines utilising Cre-LoxP and FLP-FRT systems can help target fluorophore expression to specific cells (Madisen et al., 2015; Sauer, 1987). Recombinant vector tracing tools are typically replicant disabled and also fall under the static viral tracer class as the virus dependent labelling cannot spread from the initially infected cells. Anterograde labelling is achieved by tracer binding to receptors on the soma with directional propagation of the virus to cell processes. Retrograde labelling is achieved by tracer binding to receptors preferentially expressed at axonal terminals with virus propagation to the soma following internalisation. Functional differentiation between soma, dendrites, and axon gives neurons polar physiology. Selectivity between anterograde and retrograde labelling is, therefore, important for understanding information flow in circuits. Co-injection (i.e. simultaneous anterograde and retrograde injection) can be powerful for swiftly dissecting input-output pathways in nuclei of interest (Thompson & Swanson, 2010; Zingg et al., 2014). Viral vectors are derived from many constructs, including the human immunodeficiency virus (HIV) and the herpes simplex virus type-1 (HSV-1). However, the adeno-associated virus (AAV) is the most widely used construct due to lower toxicity, rapid uptake, and rapid transgene expression within two weeks. Directed evolution of the AAV construct has also lead to efficient retrograde labelling variants (Tervo et al., 2016).

Static labelling is effective for mapping soma and projections resident in a target location of the brain. However, sometimes we desire demonstrative interrogation of functional connectivity requiring the labelling of connected neurons (Saleeba et al., 2019). Engineered wild-type neurotropic viruses, such as the rabies virus (RV; Kelly & Strick, 2000) and the

pseudo-rabies virus (PRV; Card et al., 1990), can be modified to become conditionally transsynaptic which provides controlled infection of the central nervous system. For example, glycoprotein (G-protein) deleted rabies virus (RVdG; Callaway & Luo, 2015; Wickersham et al., 2007) is only capable of transsynaptic spread if G-protein is supplied to the initially rabies infected neuron. Additionally, pseudotyping of RVdG with avian sarcoma leucosis virus glycoprotein (EnvA) limits cellular entry to cells which express the subgroup A receptor for avian sarcoma and leukosis virus (TVA). Priming the starter population of neurons for RV infection is achieved by delivering a transcription gene encoding for the TVA receptor, G-protein, and a fluorescent protein for tagging. Once TVA receptors and G-protein are expressed, RVdG pseudotyped with EnvA can be delivered to the same location where binding of EnvA to TVA receptors allows RV entry to the cell. G-protein present in primed neurons permits RV to spread transsynaptically. Strains of RV exhibit efficient and exclusive retrograde transport (Kelly & Strick, 2000; Ugolini, 2011) meaning transsynaptically infected neurons constitute monosynaptic pre-synaptic neurons (Callaway & Luo, 2015; Wickersham et al., 2007). Lack of G-protein in infected pre-synaptic neurons, assuming the pre-synaptic neuron is not itself part of the starter population, prevents further spread of the virus.

Whilst immensely powerful as a tool to investigate neuronal connectivity, debate and ongoing research surrounds the underlying mechanism and validation of the types of neurons which are susceptible for transsynaptic infection (Sun et al., 2019). Furthermore, rabies virus infection leads to neuronal death (Fu & Jackson, 2005; Wickersham et al., 2007) motivating recent exploration of different RV strains which show various degrees of neurotoxicity and efficacy (Callaway, 2008; T. R. Reardon et al., 2016). Furthermore, synapse specificity of RV infection is not well documented (Wickersham et al., 2007). Despite this, RV still provides an exceptional tool that provides a basis for studying functional pathways and was therefore explored in this thesis to help visualise input pathways onto target cell types. Recent exploration of AAV variants suggests effective anterograde transsynaptic labelling, which opens the door to downstream connectivity labelling (Zingg et al., 2017). However, the use of engineered AAV variants is likely to be met with the same levels of scepticism seen with the neurotropic viruses. Evidently, further comprehensive studies are required to ensure techniques are non-biased.



## 1.7 RESEARCH FOCUS

The topics discussed above cover biological, biomedical and computational engineering aspects of the research portrayed in this thesis. Here, we discuss the scientific focus of the thesis and how the techniques above can be used to address several scientific topics. The first scientific area addresses the visual circuitry system and the connectivity of local interneurons found in the dorsal lateral geniculate nucleus (LGd) of the mouse. The second scientific area of research is the age-related changes in cognitive performance in mice and the associated anatomical correlates.

The thalamus plays a vital role in sensory information relay to the cortex. In the visual system, afferent projections of retinal ganglion cells (RGCs) primarily innervate the first-order lateral geniculate nucleus (LG) despite there being several retinorecipient targets in the rodent thalamus (Monavarfeshani et al., 2017). The LG is split into three components; the dorsal lateral geniculate nucleus (LGd), the intergeniculate leaflet (IGL), and the ventral lateral geniculate nucleus (LGv). The LGd establishes the primary relay nucleus for image-forming visual information to the primary visual cortex (VISp) and exhibits specific cytoarchitecture in the rodent mouse with eye-specific retinorecipient domains (Godement et al., 1984) and RGC class-specific laminae retinorecipient arrangement (Kerschensteiner & Guido, 2017). Thalamocortical (TC) cells are the most abundant cell type in the LGd at around 90% of the neuronal population and cluster into three classes; W, X and Y-like cells (Krahe et al., 2011). Each TC class is distinctly distributed across the LGd with W-like cells found in the shell and X and Y-like cells found in the core (Krahe et al., 2011). Local RGC arrangement in the retina is also spatially preserved in the LGd by having RGC afferents exciting spatially corresponding TC cells through large and often complex terminal arrangements (Bickford et al., 2010; Sherman, 2004). TC cells relay visual information to VISp layers 1 and 4 and also receive modulatory cortical feedback through small terminal inputs originating from VISp layer 6 projection neurons (Sherman, 2016).

Other TC inputs are inhibitory and arise from the local GABAergic interneuron population, which form the remaining 10% of the neuronal population in the LGd (Arcelli et al., 1997; Evangelio et al., 2018), the GABAergic reticular nucleus (RT) and also the pretectum (Born & Schmidt, 2007). In rodents, the local interneurons are absent from almost all other dorsal

thalamic regions with thoughts that this might reflect mammal complexity (Arcelli et al., 1997). Recent studies have shown that this small interneuron population inherently found in the LG can be split into two types based on morphology, gene expression and membrane property (Leist et al., 2016). Besides establishing connections with the TC cells (Jager et al., 2016; Moore et al., 2018; Morgan & Lichtman, 2020), there is little known about their functional or anatomical role in the visual thalamus relay circuitry. For example, do local LGd interneurons participate in any interneuron-interneuron activity? In addition, known corticogeniculate projections (neurons which project from primary visual cortex to the lateral geniculate nucleus) from the VISp innervate the RT en-route to the TC relay cells. Recruited RT projection neurons then inhibit the TC cells (Sherman, 2016). However, do corticogeniculate projections innervating local interneurons also target the RT en-route meaning the RT provides blanket inhibition of both the TC cells and interneurons?

Connectivity and functional mapping of the local LGd interneuron population through high-resolution imaging techniques and viral network mapping could prove insightful to add to our understanding of the complex visual thalamic system. Indeed, the extensive connectivity of the visual system across the brain makes techniques such as STP tomography well suited to this study. This further motivates the development of a novel analysis pipeline to help register whole-brain image data with a deep-learning method for automated cell detection across large volumes of image data.

Although outside the scope of this thesis, research into the visual processing system has many clinical implications. Patients with visual impairment can have complications at any stage of the visual processing pathway. However, LG or cortical prostheses can offer rehabilitative measures for those where retinal or optic nerve stimulation is not viable (Farnum & Pelled, 2020). Research into the connectivity of LGd interneurons and their distribution can therefore benefit the design and placement of stimulation devices based on the specific pathology of the patient.

The second scientific area of interest addressed connectivity changes in the brain with age and how this relates to cognitive decline. Pre- and post-natal changes in brain networks are already known to occur. In the thalamic visual system of rodents, for example, local

interneurons appear to migrate into the LGd over a two-week post-natal development period (Jager et al., 2016). However, age-related studies on brain connectivity are surprisingly lacking. Healthy ageing can lead to a reduction in many cognitive domains, including intelligence (Cattell, 1987; Salthouse, 2012) and memory (Rönnlund et al., 2005). There can also be changes in white and grey matter volume, plus the degradation of myelin through the development of white matter anisotropies. Such changes can lead to reduced reaction times and information processing speeds (Baloch et al., 2008; Salat et al., 2005; Tournier et al., 2004).

With a population of those over 65 years of age expected to increase 16% by 2050 (United Nations, 2019), there is an understandable concern of knowing when some individuals might reach severe states of cognitive decline and require daily living assistance (The King's Fund, UK). Assuredly, not everyone experiences cognitive decline at the same rate, and some even maintain high levels of cognitive ability by recruiting homologous regions of the frontal cortex (Cabeza, 2002; Cabeza et al., 2004). It is also thought that those who show high attainment levels in early life might provide themselves with a buffer for cognitive ageing in later life through a theory of cognitive reserve (Singh-Manoux et al., 2011). Unfortunately, age-related longitudinal cognition studies, such as the Lothian cohorts study (Deary et al., 2013), are impractical and time-intensive. Cross-sectional study designs offer greater study convenience but are compounded by external factors such as the environment, levels of education and employment levels. Mouse models provide a valuable asset for longitudinal ageing studies due to homogeneity in inbred mouse strains, the ability to control their environment and an endpoint of just over two years.

Behavioural experiments in mice can help measure levels of cognition (Hodges, 1996), and elucidate our understanding on any behavioural changes with age, and whether it is possible to track individual animals such that we might predict how they perform in later life. Additionally, high-resolution imaging and viral tracing techniques provide an interesting opportunity to address anatomical changes in the mouse brain with age. When combined with a behavioural study, there is an unprecedented opportunity to explore possible correlates between cognitive ageing and specific fine structure changes in the brain.

The significance to age-related research in mice is two-fold. First, this research is novel by aiming to longitudinally track cognitive behaviour over the healthy lifespan of wild type mice. Complex behavioural tasks such as the radial maze can provide many measures of various sensitivity which can be assessed for correlations and potentially allow for more robust prediction of late-life performance from early-life performance. Second, it is important to be able to establish whether cognitive measures observed in rodents is translatable to humans. Future treatments and interventions tested on rodents would therefore have more, or less, credence concerning the efficacy in humans.

In consideration of both these scientific areas, this thesis utilises a commercially available STP tomography system (Ragan et al., 2012) to conduct high-resolution imaging of the whole mouse brain (see 'Chapter 2'). We created a custom pipeline for registration of whole mouse brain data to a standardised atlas (see 'Chapter 3'). Registration results were combined with a custom automated deep-learning approach for automated cell counting in targeted brain regions (see 'Chapter 4'). RV tracing and neuronal fills with biocytin revealed interneuron morphology and connectivity in the LGd, informing us of previously unseen connectivity (see 'Chapter 5'). Lastly, we performed a longitudinal ageing study in mice using a radial maze task (Abuhamdah et al., 2016), novel object recognition task, and novel object location task. We show the ability to track individual mice based on single or multiple cognition measures. We then demonstrate how high-resolution STP tomography and whole brain MRI can help investigate age-related anatomical changes at the level of white matter and grey matter volume changes and at the level of neuronal connectivity changes in the PFC (see 'Chapter 6').

## **HIGH-RESOLUTION SERIAL TWO-PHOTON TOMOGRAPHY FOR EX-VIVO WHOLE BRAIN IMAGING**

### **ABSTRACT**

Local and distal interrogation of brain circuits requires high-resolution brain-wide imaging techniques. Serial two-photon (STP) block-face tomography offers improved spatial resolution over light-sheet techniques and improved axial resolution over confocal microscopy. Here, we modified and evaluated an STP tomography system (TissueVision Inc.) for high-resolution imaging of labelled cells, neurites and small fluorescently labelled structures across the mouse brain. We created a custom stitching pipeline which was executed alongside image acquisition to improve system throughput. We performed system and analysis pipeline evaluation using a Sox14gfp/+ transgenic mouse line and demonstrated vast thalamic and hypothalamic GFP expression. We also performed whole-brain imaging in a Methoxy-X04 stained Alzheimer's 5xFAD mouse brain which displayed extensive amyloid-beta ( $A\beta$ ) pathology at 12 months of age. Ability to resolve submicron structures, such as axons and dendrites, show tremendous capability for the system to investigate neuronal circuits.

## 2.1 INTRODUCTION

Work from the grandfather of structural biology, Ramón y Cajal, in the late 19<sup>th</sup> and early 20<sup>th</sup> century encapsulates one of the most compelling and challenging areas of neuroscience (Ramón y Cajal, 1909; Ramón Y Cajal, 1894). That is, understanding the cytoarchitecture of the central nervous system. Brain circuits function through an ensemble of functionally interconnected neurons. Anatomically, neurons consist of soma, dendrites and axon. Functionally, the dendrites and soma act as receivers for electrical signalling between neurons, whilst the axon is responsible for transmitting signals to other cells. With regards to physical size, neuronal somas are relatively small at around ten microns in diameter with larger cells, such as pyramidal cells and basket cells in the cerebral cortex, reaching 100-200  $\mu\text{m}$  in diameter (Kawaguchi & Kubota, 1998). Dendritic and axonal projections are only a few microns wide but can span local and distal volumes of the brain. For example, the cerebral cortex of the brain, which primarily processes motor, sensory and association information, contains pyramidal cells which can have axons that span multiple cortical domains and hemispheres. In one case, a single neuron was observed to wrap its projections around the outermost layers of the entire mouse brain (S. Reardon, 2017). In another case, a single neuron can have neurites spanning a total length of 6000  $\mu\text{m}$ , within a single nucleus (Morgan & Lichtman, 2020). To understand anatomical and functional connectivity, we have to track individual structures of only a few microns in diameter across the entire mouse brain.

As a general principle in biological imaging, there exists a trade-off between imaging resolution and the volume of tissue that can be imaged. Since the first magnetic resonance image was published in 1973 (Lauterbur, 1973), magnetic resonance imaging (MRI) saw rapid development. It has ultimately become a gold-standard technique employed across the world for rapid high-resolution imaging of large volumes of tissue. Modern MRI systems have an acquisition voxel size of around one cubic millimetre. This resolution is sufficient to delineate anatomical domains in larger mammalian tissue volumes such as the human brain. However, for a mouse brain with an approximate one cubic centimetre volume, the resolution offered by MRI is inadequate to robustly delineate smaller brain regions such as sub-thalamic and hypothalamic structures. Visualising individual cells is, therefore, unachievable.

Multiple properties of the magnetic resonance phenomena enable MRI to detect multiple tissue properties. T1, T2 and DTI present the three most commonly used MRI scan configurations. If looking at brain tissue, T1 is typically used to reveal fat and blood content whilst T2 is more suited for revealing fluid properties of the tissue. Meanwhile, DTI is used for measuring anisotropy and diffusivity in tissue and can reveal structural information about axon fibre tract integrity (Mori et al., 1999). Recording brain-wide activity with functional-MRI (fMRI) is also possible in awake mice performing behavioural tasks (Fonseca et al., 2020), enabling fundamental behavioural neuroscience research. However, using fMRI for brain circuit investigation can be fallacious. fMRI used to reveal neural correlates of behaviour measures the haemodynamic (blood oxygen level dependence; BOLD) response in the brain rather than the direct activity of neurons, for which the haemodynamic activity and neuronal activity have a non-linear relationship (Logothetis et al., 2001). Neuronal activity is inferred through fMRI, and statistical signal analysis can lead to misrepresentative results. Statistical misinterpretation can be illustrated by Bennett et al. (2009) who significantly showed post-mortem functional activity in a salmon during a behavioural task.

MRI is invaluable for biomedical imaging, but it is not capable of enabling brain-wide circuit mapping. Sub-cellular resolution techniques are, therefore, required. However, these higher-resolution techniques typically suffer severe limitations for imaging depths. Electron microscopy (EM) and scanning electron microscopy (SEM) can achieve a resolution around a single nanometre (Kourkoutis et al., 2012). Volume imaging necessitates serial sectioning of fixed frozen tissue. However, even the most advanced serial SEM systems would take 100 years to image a one cubic millimetre volume at a resolution of  $8 \text{ nm}^3$  (Xu et al., 2017). EM is unparalleled for use within single synapse connectivity studies (Briggman et al., 2011), but this approach is arguably excessive for mapping neuronal circuits.

Light-based optical microscopy offers impressive compromise with imaging depths down to a millimetre in a single section of tissue, whilst providing micron and submicron resolution. High-resolution light-based microscopy techniques include confocal microscopy (Minsky, 1988) as well as multiphoton techniques such as two-photon (2P) (Denk et al., 1990; Göppert-Mayer, 2009) and three-photon microscopy. However, light-based imaging methods for

biological tissue introduce challenges such as wavelength scattering and restricted imaging depth.

Autofluorescence refers to the intrinsic fluorescence of biological molecules such as haemoglobin, riboflavins, melanin and lipofuscin. These endogenous fluorophores can be excited-by and emit-within a wavelength range of 250 nm to 800 nm (Berezin & Achilefu, 2010). Intrinsic fluorescence of these molecules is highly valuable for diagnostic applications as it does not require any staining procedures (Monici, 2005). However, for imaging neuronal structures, autofluorescence can mask and reduce the signal of interest over background information, making it challenging to resolve structures. Fluorophore delivery to tissue structures either through staining protocols or injectable viral delivery techniques can help boost signal intensity and is generally the preferred approach for studies where endogenous fluorescence is not the specific topic of interest. Fluorophore design is therefore essential to ensure probes are quickly excited without eliciting emission from background molecules. Probes with blue and green excitation wavelengths are only compatible with studies which do not require high penetration into the tissue, as Rayleigh scattering in combination with high refractive index anisotropy in biological tissue leads to high laser scattering effects and ineffective laser focusing under the surface of the tissue. On the other hand, red and infra-red wavelengths scatter less in tissue, with wavelengths longer than 800 nm providing an ideal excitation regime as endogenous fluorescence is lowest (Kobayashi et al., 2010). Wavelengths between 800 nm and 1100 nm are considered to be within the optical window, as those wavelengths exhibit reduced scattering effects and reduced excitation of intrinsically fluorescent molecules.

A confocal microscope can achieve a spatial resolution of 0.2  $\mu\text{m}$ . However, imaging depth is limited to 200  $\mu\text{m}$  to 250  $\mu\text{m}$  before laser scattering becomes significant, and the detection of out-of-focal-plane fluorescence limits the resolving of small structures (So et al., 1998). Confocal microscopy also illuminates an axially distributed double-conic volume of tissue above and below the focal point, leading to phototoxic damage to tissue or fluorophores in regions which have not been imaged yet. The laser requires point scanning over the objective field-of-view (FOV) using high-speed scanning galvanometer mirrors, with emission light



detected using photomultiplier tubes. Confocal imaging is therefore much more time-intensive than camera-based detection system as with widefield imaging techniques.

2P microscopy is a compelling alternative to confocal microscopy and offers several benefits which make it suited for imaging biological samples. Introduced by Denk et al. (1990), 2P microscopy relies on the co-absorption of two photons at an axially and spatially confined focal point around  $0.1 \mu\text{m}^3$  (Zipfel et al., 2003) with each photon carrying half the energy (i.e. double the wavelength) required for fluorescence excitation. Photon-photon interaction has a quadratic relationship to the density of photons, meaning high temporal and spatial incidence of photons is required at the focal point, using high powered femtosecond mode-locked lasers. The longer wavelength of the excitation beam means 2P microscopy exhibits reduced scattering effects enabling deeper imaging down to routine depths of around 400 to 450  $\mu\text{m}$  (Durr et al., 2011; Takasaki et al., 2020). In one case, an *in-vivo* set up demonstrated imaging to a depth of 1.6 mm (Kobat et al., 2011). The excitation wavelength in a 2P configuration shifts towards the optical window for a more optimal reduction in background fluorescence stimulation. Axial and spatial confinement of the excitation beam also reduces photobleaching and phototoxicity everywhere except the focal point of interest. The drawback to 2P microscopes is the requirement to point scan the focal volume across the FOV, making acquisition slower.

Light-sheet microscopy (LSM) offers an alternative high-resolution imaging method with faster acquisition and capability to image volumes of tissue (P. J. Keller & Dodt, 2012). Light sheets, generated with a cylindrical lens, illuminate a planar section of the sample and an orthogonally positioned objective collects emission (Girkin & Carvalho, 2018). Illumination of an entire FOV enables fast camera-based detection. The light sheet, or the sample, can be moved over the sample for rapid excitation and acquisition. For example, it takes just hours to image the mouse brain (Renier et al., 2014). A single plane is also excited at any one time reducing phototoxicity in the axial direction. However, planar illumination in the spatial domain would still cause phototoxicity in the spatial dimension. A caveat with LSM comes with the ability to transmit light through an entire sample effectively. Residual light scattering in samples means LSM generally has reduced contrast in comparison to confocal or 2P microscopy and can lead to shadowing artefacts. Simultaneous illumination using two

opposing light sheets improve contrast, reduces shadowing, and also leads to a more uniform illumination profile (Huisken & Stainier, 2007). Confocal light-sheet microscopy (CLSM) offers vastly superior contrast imaging over standard LSM and has been used to image Purkinje cell soma and neuronal projections (Silvestri et al., 2012).

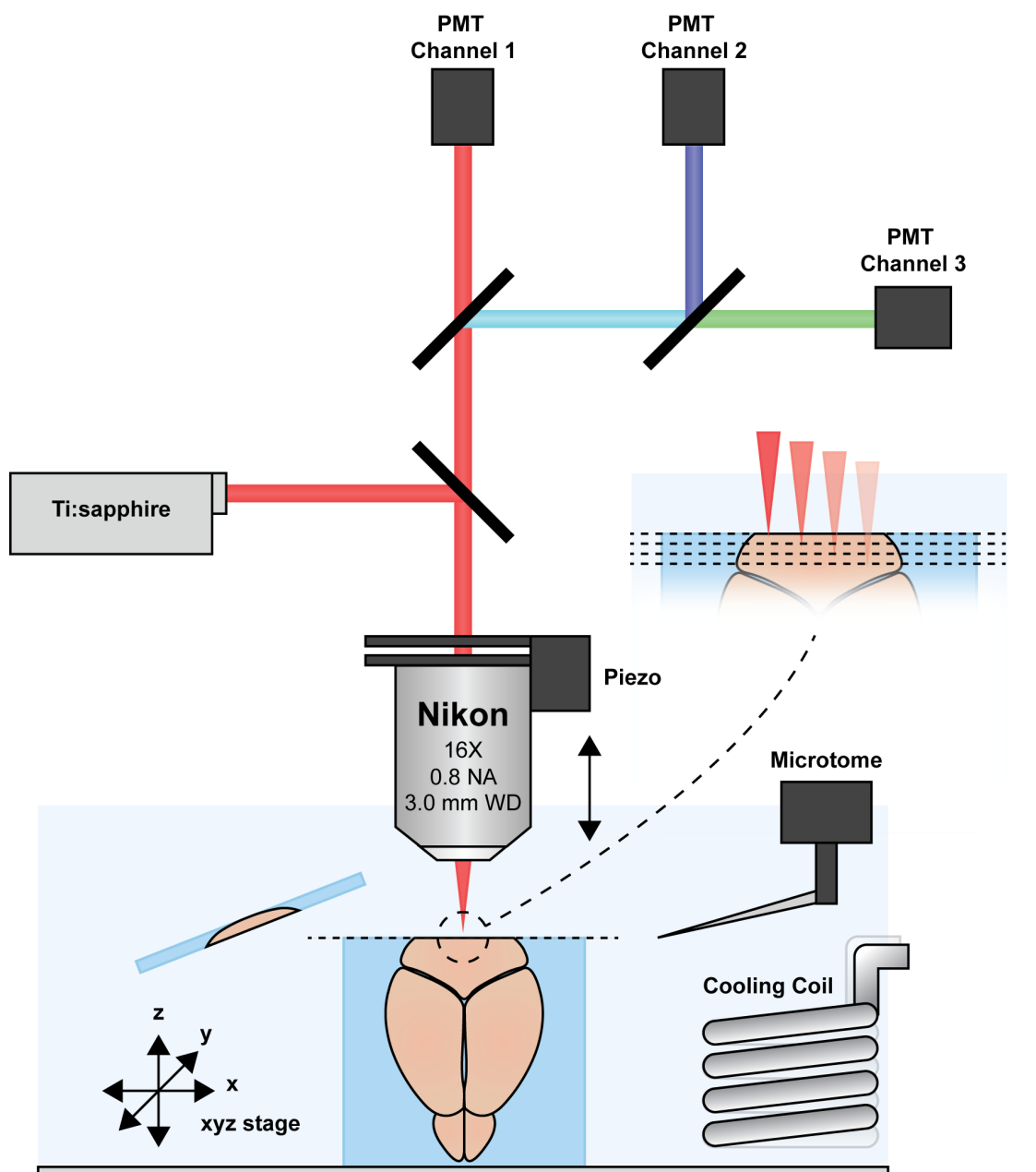
However, LSM and CLSM on mouse brain tissue are only possible when combined with methods to render tissue transparent by reducing the variation in refractive index between tissue structures (Silvestri et al., 2016). Improved light transmission through the tissue can be accomplished by using organic solvents to replace water content with a high refractive index medium, used within BABB or DISCO based techniques, for example (Ertürk et al., 2012; Renier et al., 2014). Aqueous solutions with a high refractive index such as sucrose, fructose and FocusClear provide effective clearing methods through simple immersion techniques and raise the refractive index to that of the lipids. Reducing the refractive index of proteins is achievable through hyper-hydrating solutions such as in CUBIC (Susaki et al., 2015). Lastly, a hydrogel polymer can be used scaffold tissue structure before using a detergent to strip away lipids as presented in CLARITY (Chung et al., 2013). Clearing a mouse brain can enable light-sheet imaging at the sub-cellular resolution, but there are many caveats associated with clearing. Stripping away lipids or dehydrating tissue can result in anisotropic shrinkage, which could affect fine-structure connectivity (Silvestri et al., 2016). Clearing protocols can also take days to accomplish and may only work on some types of tissue and tissue size. They also might not be compatible with primary fluorescence from fluorophores (Renier et al., 2014). In some clearing protocols such as the solvent-based methods, where the refractive index of tissue uniformly raised, imaging requires immersing the sample in a high refractive index matching solution such as BABB (Ertürk et al., 2012). These solutions can be expensive and require specialised imaging systems and objectives which are compatible with immersion solutions. A high numerical aperture objective with a long working distance is required for cleared samples which remain intact during imaging.

Overall, if it is not possible and practical to use tissue clearing procedures because of the adverse effects of clearing or lack of compatible imaging system, stereotaxic sectioning provides the only viable alternative. Individual and manual sectioning techniques are labour intensive and leads to tissue distortion and alignment issues. These issues can affect

downstream analysis which might require registration algorithms to register multiple sections. In some cases, distortion might be severe to the point of data becoming unusable. The alternative to this comes in the form of an automated block-face imaging system. These systems prepare tissue samples in a block, such as an agarose block, which is imaged by serial sectioning and scanning over with an imaging objective. Such a system is available from TissueVision Inc., called TissueCyte 1000 (Ragan et al., 2012), which combines serial sectioning with a 2P microscope to permit high-resolution imaging across large volumes of tissue. Using serial two-photon (STP) tomography in combination with axially confined optical sectioning, reduces the amount of physical sectioning, thereby reducing acquisition time.

Briefly, STP tomography with TissueCyte 1000 works in the following way (see 'Methods' for a more detailed description; Fig. 2.1). We embed a sample in agarose and fasten it into a water bath mounted onto a motorised XYZ stage. We configure imaging parameters using the provided software suite, including key parameters such as defining the volume in XY and Z to be imaged, and configure the slice thickness for serial sectioning. During imaging, the sample is translated underneath the objective to acquire a tiled FOV as either a single planar image or as an optical Z-stack over a defined depth. The XYZ stage moves the sample in the XY plane until the tile FOVs cover the XY imaging area. Once all tiled positions are acquired, a mounted microtome sections the tissue from the surface of the agarose block. The slice remains free-floating in the bath and the imaging and sectioning procedures are repeated until the volume has been imaged.

Here, we evaluated TissueCyte 1000 for high-resolution imaging across the entire mouse brain, such as to assess its ability to facilitate studies on brain-wide connectivity. We made several modifications to the system to improve the imaging conditions and throughput as described under 'Methods'. Furthermore, we engineered an image stitching pipeline to help with the initial image processing of the generated data. We tested the system by imaging the amyloid- $\beta$  (A $\beta$ ) plaque distribution across a Methoxy-X04 stained 5xFAD Alzheimer's mouse brain. Next, we imaged the brain from a Sox14gfp/+ transgenic mouse to perform single-cell resolution mapping of the Sox14 neuronal population. Finally, we performed a co-injection of anterograde pAAV-CAG-tdTomato and retrograde pAAV-hSyn-EGFP labelling in the dorsal lateral geniculate nucleus (LGd) to evaluate viral labelling use with the system.



**Figure 2.1.** *TissueCyte 1000, for serial-two photon imaging across tissue volumes. Samples are embedded in agarose, fastened to a motorised XYZ stage and submerged under phosphate buffer. A microtome attached to the side of the stage enables sectioning whilst the 16× immersion objective images the sample using a Ti:sapphire (Chameleon Ultra II; Coherent) with emission collected across three PMTs for red, green and blue channels. An adapted cooling coil helps keep the sample and free-floating slices at 4°C for tissue preservation for further histological analysis. A piezo control on the objective allows additional optical sectioning.*

## **2.2 METHODS**

### **Plaque labelling in a 5xFAD mouse**

A female 5xFAD (12 months) mouse was intraperitoneal (IP) injected with 10 mg/kg of Methoxy-X04. After two days, the mouse was anaesthetised with intraperitoneal (IP) injection of sodium pentobarbital (50 mg/kg) and transcardially perfused with 20 ml of 1X phosphate-buffered saline (PBS, pH 7.4) at 4°C, followed by 20 ml of 4% paraformaldehyde (PFA) at 4°C. After extraction, brain tissue was stored in 4% PFA at 4°C for 24 hours before washing three times with 1X PBS and storing in 1X PBS with 0.01% NaN<sub>3</sub> until imaging.

### **Sox14 expression mapping**

The brain from a Sox14GFP/+ (P21) mouse was extracted by anaesthetisation with sodium pentobarbital (50 mg/kg) and a transcardial perfusion process using 20 ml 1X PBS at 4°C, followed by 20 ml 4% PFA at 4°C. Brain tissues were post-fixed in 4% PFA and kept at 4°C for 24 hours before transferring to 1X PBS with 0.01% NaN<sub>3</sub>.

### **Co-injection for anterograde and retrograde labelling of thalamic connectivity**

Following anaesthetisation with isoflurane, an adult C57BL/6 mouse was head-fixed within a stereotaxic injection setup. The isoflurane flow rate was modified to keep a stable respiratory rate. The injection coordinates for the LGd were AP 2.1 mm, ML 2.0 mm and DV 2.3 mm. We used a cranial drill to thin the skull and glass pipette filled with a mixture of 0.25 µl of pAAV-CAG-tdTomato (anterograde) and 0.25 µl of pAAV-hSyn-EGFP (retrograde). A co-injection was done to reduce the number of injections ensure both viruses were delivered to the same site. After 10 minutes, the glass pipette was slowly retracted. Mice were left to recover for two weeks before IP injection of sodium pentobarbital (50 mg/kg) and transcardial perfusion with 20 ml of 1X PBS (pH 7.4) at 4°C, followed by 20 ml of 4% PFA at 4°C. After extraction, brain tissue was stored in 4% PFA at 4°C for 24 hours before washing three times with 1X PBS and storing in 1X PBS with 0.01% NaN<sub>3</sub> until imaging.

### **Sample preparation for serial two-photon imaging**

Samples were embedded in agarose before STP imaging. A 4.5% oxidised-agarose solution was prepared by stirring agarose (type 1; Sigma) with 10 mM NaIO<sub>4</sub> (Sigma) and 50 mM phosphate buffer (PB) at room temperature in a light-shielded environment for 2 hr 30 mins.

The oxidised-agarose solution was filtered three times with PB before re-suspending in PB and stored at 4°C until performing the embedding process. During embedding, two 25G needles were horizontally inserted approximately 5 mm apart, through a disposable cube-shaped embedding mould (VWR #E6032-1CS) to provide a horizontal platform to support the mouse brain. The oxidised-agarose solution was heated to boiling point in a microwave and left to cool to 65°C. After filling the mould with agarose, the mouse brain was placed onto the needle platform using a set of forceps. The needle platform aided alignment of the sample within in the agarose preparation. Furthermore, slicing a flat surface from the sample was not required, meaning the entire sample could be imaged. Finally, the embedded sample was left to passively cool to room temperature until becoming solid, before carefully extracting from the moulds and immersed and stored within 1X PBS with 0.01% NaN<sub>3</sub> at 4°C until imaging.

### **TissueCyte 1000 for serial two-photon tomography**

STP tomography of an entire mouse brain was done using a TissueCyte 1000 imaging system (Fig. 2.1). Brain samples embedded in the agarose block were glued to a magnetic slide using a cyanoacrylate-adhesive (Loctite Ltd.) with the anterior-posterior axis aligned with the imaging objective. The slide was fastened into a custom-designed 1.5 L water bath using magnets to keep the sample stationary. The sample was aligned with the medial-lateral axis of the brain parallel to the X-axis of the stage, and the dorsal-ventral axis of the sample aligned with the Y-axis of the stage. The water bath was filled with phosphate buffer (PB, pH 7.4) with 0.01% NaN<sub>3</sub> to keep the sample hydrated during imaging and to prevent bacterial growth. The water bath containing the sample was fastened onto a motorised XYZ stage, and a custom cooling coil (Corio CD-200F; JULABO UK Ltd.) was positioned into the bath and set to 4°C (see the modifications below). A high-powered femtosecond Ti:sapphire (Chameleon Ultra II; Coherent) laser generated the laser excitation beam with a 690-1080 nm tuning spectral range. A 16× water immersion objective (0.8 NA, 3 mm working distance; Nikon) was attached to a piezo motor with a 400 µm range to allow independent Z-axis manipulation of the objective of optical Z-stacks without requiring the entire XYZ stage to be manipulated. The fluorescence emission path consisted of three PMTs for three-channel fluorescence collection. Channels were split into red wavelengths (>560 nm) using a Dichroic T560LPXR low-pass mirror, and subsequently splitting green wavelengths (500-560 nm) and blue (<500 nm) wavelengths using a Dichroic 500DCXR mirror. A microtome mounted to the side of the

water bath provided physical sectioning. The microtome blade extended to a depth of 3 mm to be in approximate alignment with the working distance of the objective. The 400  $\mu\text{m}$  piezo control Z-axis range of the objective provided moderate leniency in this alignment.

We used the following approach for system alignment. The sample was secured within the water bath and submerged under PB. The upper surface of the agarose embedded sample was then roughly aligned with the imaging plane of the objective and the microtome blade. Next, we defined a tile imaging area in X and Y by determining the stage positions corresponding to the left-most, right-most, top-most and bottom-most extents of the sample. All positions were marked with respect to the objective focal point. The area was then divided by dimensions of a single tile to get the required number of tiles in X and Y to cover the sample. For the 16 $\times$  objective used here, a tile approximated to a 1  $\times$  1 mm FOV corresponding to 10  $\times$  12 tiles covering the coronal view. The total depth of imaging was defined by using the section thickness and specifying the number of physical sections that were required. It was possible to physically section at 100  $\mu\text{m}$  thickness, meaning the entire mouse brain is imaged in 150 to 200 sections. Typically, total imaging depth was less than this because some regions of the brain might not be relevant to the study; the cerebellum and olfactory bulb areas could be omitted, for example. The piezo control of the objective enabled optical sectioning and Z-stacking at each tile position across a 400  $\mu\text{m}$  range. In theory, optical sections could be sampled at any depth, meaning the entire mouse brain could be sampled in the Z-axis every micron. In practice, this would be excessive for most applications and would generate an extraordinary amount of data and would take several weeks to accomplish imaging of an entire mouse brain. Following Nyquist theorem, a minimum sampling every 25  $\mu\text{m}$  would sample individual cells twice. To improve the sampling, optical sectioning every 10  $\mu\text{m}$  was typically performed. During the alignment steps here, the XYZ stage would often not correctly position itself in the Z-axis. To ensure the stage was correctly calibrated, dry-run sectioning was performed by repeatedly running the sectioning process with a blade vibration frequency of 60 Hz and velocity 0.35 mm/s, as implemented during the automated acquisition process.

Once alignment and consistent tissue sectioning was confirmed, we tuned the laser excitation wavelength and set the PMTs to an ideal gain. Imaging parameters were calibrated by running

a continuous image acquisition process and modifying the objective height, laser wavelength, PMT gain, image resolution, laser scanning parameters, and the laser power adjustment if taking optical Z-stacks. Further dry-run tissue sectioning with the microtome was needed to validate consistent sectioning before initiating the STP tomography scan in autonomous mode.

During the autonomous scanning, the XYZ stage translates the sample underneath the objective at the front-left tile position where imaging for a section always starts. The laser shutter opens for 2P excitation, with emission collected across the three PMTs. All optical planes were sampled at a single tile position before translating the sample to the next tile position, if optically sectioning. Section imaging was completed once all tile positions and optical sections were acquired. The sample is then translated to the microtome for physical sectioning of tissue. Tissue slices remain free-floating in the PB solution but were not likely to drift underneath the imaging objective. The sample is returned to imaging objective to repeat the imaging procedure. Following STP scan completion, free-floating slices were extracted from the PB and stored in 1X PBS with 0.01% NaN<sub>3</sub> at 4°C. When imaging high-resolution at 2080 × 2080 pixels per tile, a scan across a mouse brain at resolution 0.54 × 0.54 × 10 μm would complete after five days of acquisition time and yields roughly 3 TB of raw data (3 MB per tile). All image data was saved onto a network-attached storage (NAS) drive (DS1815+; Synology Inc.).

For Methoxy-X04 imaging, we used 740 nm excitation with emission collected over the blue channel using an additional 435/40 nm BrightLine bandpass filter (Laser 2000 Ltd.) to limit signal bleed-through from the green channel. A 920 nm laser excitation wavelength was used for GFP imaging in the Sox14gfp/+ mouse, with GFP collection on the green channel with an additional 525/40 nm Brightline bandpass filter (Laser 2000 Ltd.). For pAAV-CAG-tdTomato and pAAV-hSyn-EGFP imaging, we used 980 nm excitation which provided an optimum single wavelength excitation for both tdTomato and EGFP two-photon cross-section absorption (Drobizhev et al., 2011).



## **Modifications to TissueCyte 1000**

The TissueCyte 1000 system was designed to be a complete pre-configured system, but has many shortcomings which prevent this being the case. In its initial configuration, the system was capable of robustly sectioning 100  $\mu\text{m}$ . However, thinner sectioning can be advantageous for reducing laser scattering and power attenuation at deeper tissue depths. Below 100  $\mu\text{m}$  the system was not robust at slicing and caused some slices to be obliquely acquired. Irregular slicing resulted in an uneven thickness of the tissue between the plane of imaging and the objective and led to uneven illumination artefacts. Microtome slicing issues were attributed to a steep angle-of-attack for the blade, which would snag on tissue during the sectioning procedure. In severe cases, soft tissue areas such as the caudate putamen were completely excised from the sample. The first modification to the system was the removal of the microtome and changing the blade angle-of-attack to a shallower approach. This improved slicing quality and permitted robust slicing down to 40  $\mu\text{m}$  thickness. Sectioning thinner than 40  $\mu\text{m}$  did eventually result in oblique slicing. A diamond blade or improved microtome configuration might reduce robust slicing to 20-30  $\mu\text{m}$ .

The second modification ensured the system retained an optimum imaging environment during scans which could last up to a week. During extended STP scans, the PB temperature increased from 4°C to 28°C. This raised two critical concerns. First, the PB would evaporate during long imaging sessions over a day. If the level dropped below the height of the objective, it would affect the laser excitation and emission path. If it dropped below the surface of the block, the sample would dehydrate. The second concern was that free-floating slices would degrade when left at 28°C in the PB solution. Tissue degradation resulted in the generation of extraneous tissue debris which would float between the objective and the sample and cause shadowing artefacts during image acquisition. This was resolved by designing a custom cooling coil that was attached to a commercial cooling system (Corio CD-200F; JULABO UK Ltd.). To accommodate the cooling coil, we designed a custom 1.5 L plexiglass water bath. The custom system permitted cooling of the PB to a constant 4°C throughout extended periods of imaging and prevented tissue from degrading. Cooling the slices also permitted tissue sections to be extracted after imaging for staining and histological procedures, without worrying about tissue degradation.

The third modification was to improve the spectral discrimination between all three channels. Natively, the Dichroic T560LPXR and Dichroic 500DCXR mirrors split wavelengths into the red channel ( $>560$  nm), green channel (500-560 nm) and the blue channel ( $<500$  nm). Restriction of fluorophore emission bleed-through across detection channels was achieved using a 525/40 nm Brightline bandpass filter (Laser 2000 Ltd.) installed onto the green channel PMT, and a 600/37 nm Brightline bandpass filter (Laser 2000 Ltd.) installed onto the red channel PMT.

### **Stitching pipeline**

The native stitching pipeline supplied by TissueVision required stitching in MATLAB (Mathworks) and ImageJ/Fiji (Schindelin et al., 2012) using a deprecated stitching plugin. Stitching was only possible after imaging was completed and took several days to complete. With native stitching limiting throughput, a faster custom stitching pipeline for execution alongside TissueCyte 1000 acquisition was created and is described as follows (Fig. 2.2).

Individual tiles at  $2080 \times 2080$  pixel resolution acquired through STP imaging were stitched together with a custom pipeline written in Python that calls ImageJ/Fiji. First, to expedite the stitching configuration, an output file from TissueCyte 1000 is automatically read for parameter values dictating the number of tiles to expect for stitching, the number of sections, the resolution of imaging and where the data is stored. Once the system saved all tiles per physical section, each tile was read from the network storage device onto the local workstation. Saving the tile data locally mitigated increased image read times over the network. Each tile was cropped to  $2024 \times 2024$  pixels from an original  $2080 \times 2080$  pixel size to remove artefacts on the edges of each tile. A previously acquired average flatfield image was gaussian blurred (kernel size  $201 \times 201$  pixels), divided by its maximum intensity value, and subtracted away from integer value 2. This converted the flatfield image into a correction image. When opting not to use the flatfield image, an approximation can be generated from 1000 randomly sampled non-blank tile images. We applied an intensity ramp defined by a quarter-circle to a 144 pixel-wide boundary along each edge of the correction image, resulting in a non-linear reduction in pixel intensity to zero value at the absolute edge of each tile. Intensity ramping reduced tiling artefacts when tile images were stitched together (Fig. 2.2B). Each STP tile was multiplied by the intensity-ramped correction tile and stitched using the

Grid/Collection stitching plugin in ImageJ/Fiji using maximum intensity projection along the overlapping regions. Stitched images were converted to 8-bit by appropriately scaling the 16-bit image range, and additionally downsized to an isotropic voxel size of 10  $\mu\text{m}$  for post hoc computational analysis.

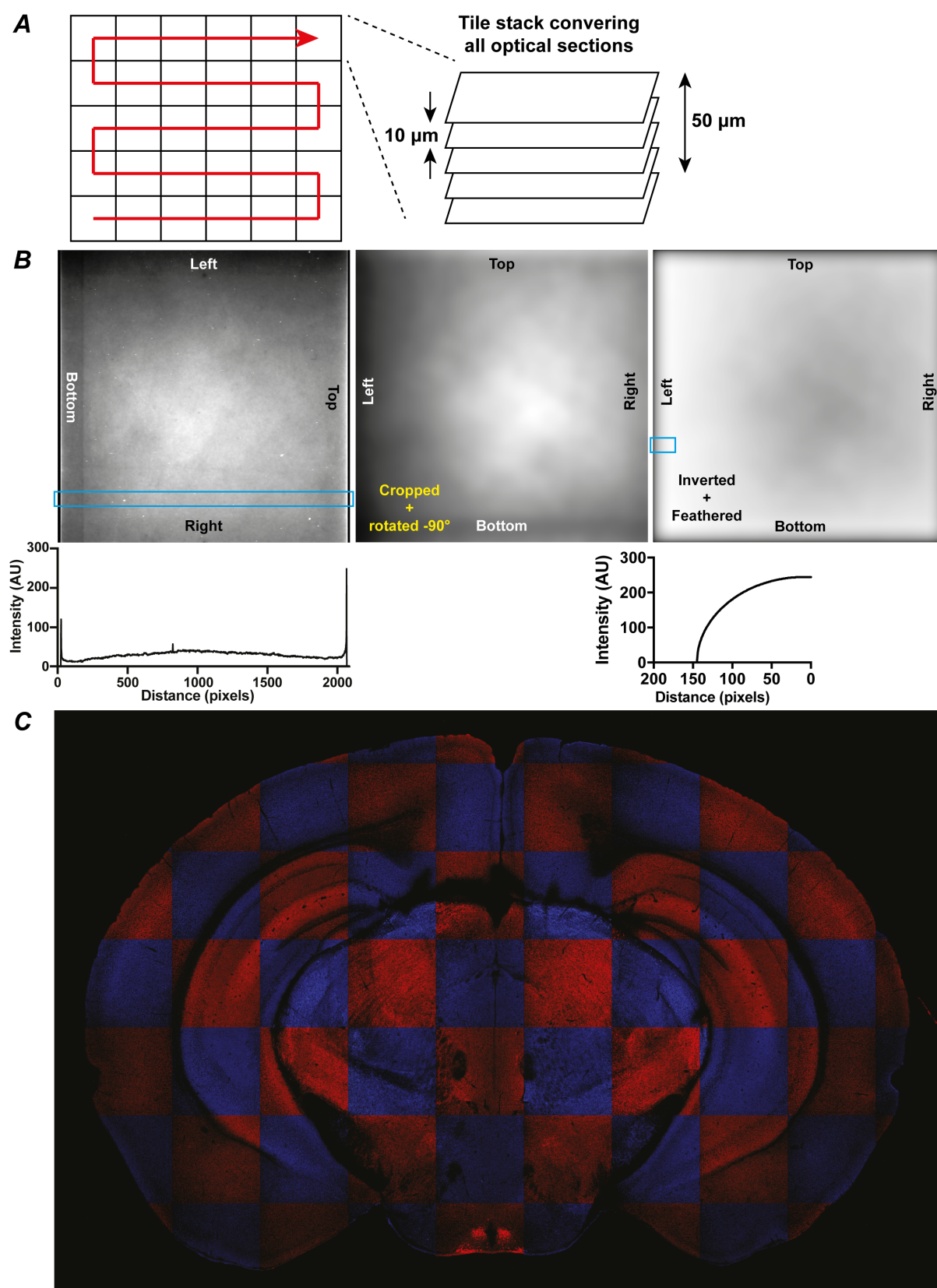
### **Visualising data**

To visualise STP data, ImageJ/Fiji was used to load individual and multiple stitched coronal sections. Depth-coded stacks for unique colour mapping each Z position in a Z-stack was done using the Cookbook plugins in ImageJ/Fiji. Visualisation of whole-brain data was done by downsampling the total Z-stack to  $5 \times 5 \times 10 \mu\text{m}$  voxel size and viewing in Imaris (Bitplane; Oxford Instruments).

## **2.3 RESULTS**

### **Improved tile stitching process**

The improved tile stitching pipeline was executed alongside STP acquisition and drastically expedited the stitching procedure. STP imaging of a 50  $\mu\text{m}$  thick section thickness with 10  $\mu\text{m}$  optical sections completed in 30 minutes (Fig. 2.2A). Stitching all tiles for a single optical section was completed in around 2 minutes, with all stitching completed 10 minutes after the STP scan concluded (5 optical sections  $\times$  2 minutes per tile). This stitching pipeline resulted in a drastic improvement in processing time compared to the original pipeline, which only started after imaging was completed and took days to process. Stitching in parallel with image STP acquisition also afforded the ability to halt and modify the scan configuration, which proved valuable for longer scans. Generating a correction tile (Fig. 2.2B) and feathering using a circular function on the overlap edges of the tiles improved the overall quality of the final stitched image by reducing edge-artefacts in the tiles that could propagate into the stitched result. Tile stitching was a necessary first step in processing in order to generate a complete coronal section of tissue for viewing and downstream analysis methods (Fig. 2.2C).



**Figure 2.2.** Image stitching pipeline. **A.** TissueCyte 1000 acquires tiles in a snaking pattern from the front-left tile position. All optical sections are acquired per tile position as piezo control of the objective in the Z-axis is faster than the translational movement of the motorised

stage in the Z-axis. **B.** A flatfield image is cropped around the edges to remove high-intensity artefacts. Photobleaching on the trailing edges of tiles (left and bottom edges) results in a reduced intensity, as shown in the intensity plot of the indicated region in blue (left). A Gaussian filter is applied to the flatfield tile to smoothen out the intensity distribution and rotated 90° to orientate the tile with the imaging axis (middle). The flatfield tile is subtracted from integer value 2 to give a multiplicative correction tile and feathered along the edges with a circular function (right). The final tile is multiplied against all tiles in the section currently being stitched to correct for uneven illumination and improve the final stitched result. **C.** All tiles per optical section are fed into ImageJ/Fiji and stitched with the Grid/Collection stitching plugin. Individual tiles are visualised here with alternating blue and red colour maps. A single stitched section of 12 × 10 tiles takes around 2 minutes to complete.

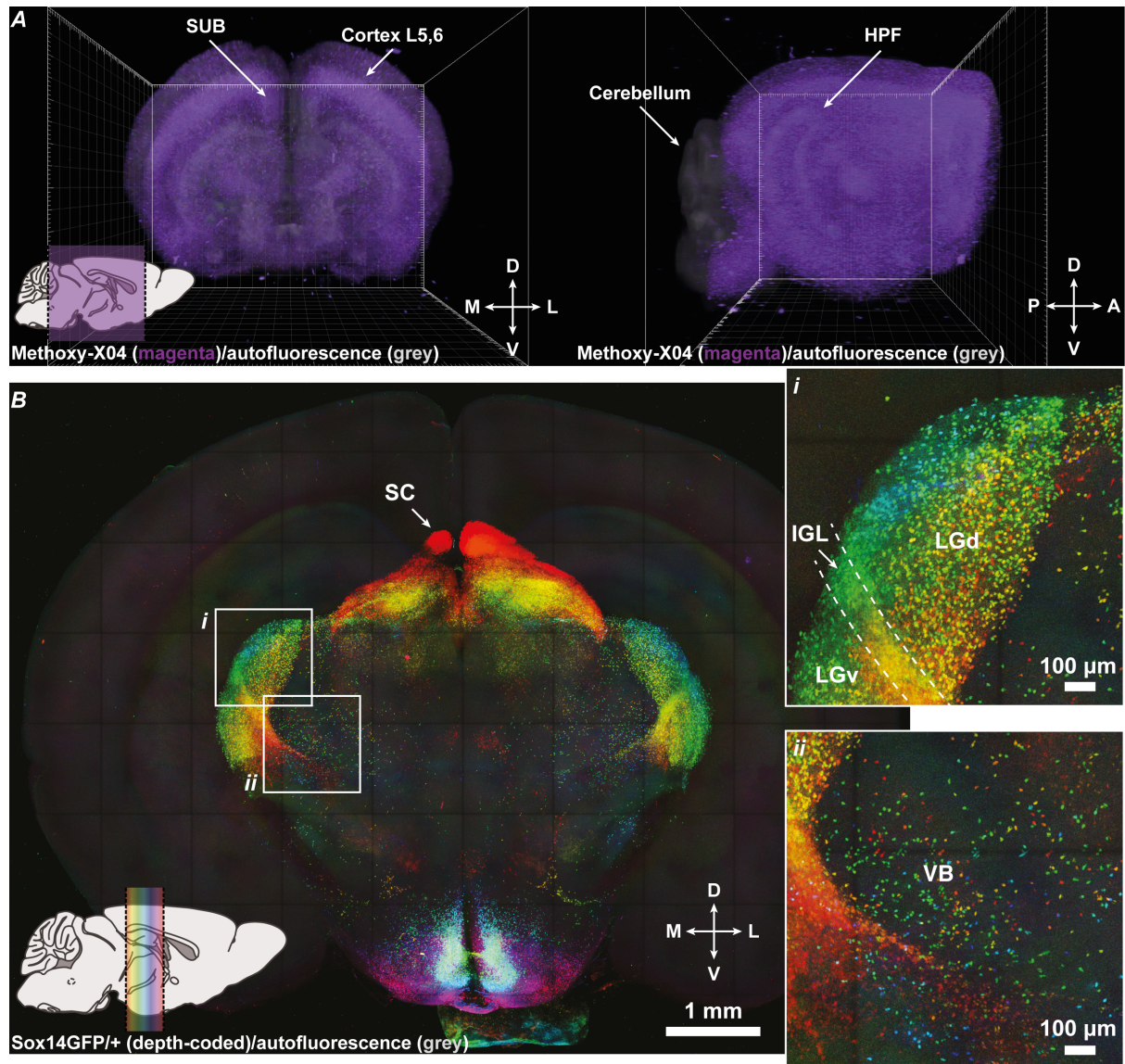
### **STP imaging of Aβ plaques shows aggressive brain-wide amyloid pathology**

STP imaging Methoxy-X04 stained Aβ plaques in a 12-month-old 5xFAD mouse shows an impressive capacity to map plaque distribution throughout a large volume of the mouse brain (AP +0.4 to AP -6.6; Fig. 2.3A). An acquisition pixel resolution of 0.54 μm allowed individual plaques to be identified. Additionally, multichannel imaging of Methoxy-X04 alongside background autofluorescence enabled segmentation of Aβ plaques over the background signal. Spectral delineation between signals offers a starting point for deeper imaging analysis. In the exemplar demonstration here, the 5xFAD mouse model represents an aggressive Alzheimer's model with Aβ plaques most prominently located in the isocortex and predominantly layers 5 and 6. The subiculum and hippocampus formations (HPF) also show a marked increase in amyloid pathology over neighbouring areas. The cerebellum shows the least pathology compared to the rest of the brain volume. This result is in agreement with known pathology for the 5xFAD mouse model which exhibits severe subiculum and layer 5 Aβ deposits from as young as 16 days (Richard et al., 2015) with neuronal loss observed in layer 5 pyramidal neurons after 12 months (Jawhar et al., 2012). In theory, a neuronal stain could quantify neuronal loss using the complete STP imaging approach demonstrated here.

### **STP imaging of the Sox14gfp/+ mouse shows an extensive thalamic and hypothalamic distribution of Sox14 gene expression**

High-resolution imaging in the P21 Sox14gfp/+ mouse revealed the widespread thalamic distribution of Sox14 expressing cells across first and higher-order sensory nuclei, not previously reported (Fig. 2.3B). High gene expression observed in the dorsal lateral geniculate nucleus (LGd) agrees with the co-localisation of Sox14 expression with the local GABAergic interneurons in the LGd (Jager et al., 2016). The superior colliculus (SC) showed extensive GFP expression, resulting in saturation of the PMTs and inability to distinguish between individual nuclei (Fig. 2.3B). The high SC GFP expression allowed the Sox14 transgenic mouse line to being phenotyped by looking at the SC under a widefield fluorescence microscope. High spatial resolution of 0.54  $\mu\text{m}$  allowed clear distinction between individual cells. However, clear distinction was not possible in regions where there was sufficiently dense aggregation of fluorescent signal such as in the SC and periventricular regions of the hypothalamus (Fig. 2.3Bi-ii). High spatial-resolution combined with strong fluorescence morphologies opens the door for downstream analysis methods, such as cell counts and density quantification.





**Figure 2.3.** Exemplar results for serial two-photon whole-brain imaging and stitching. **A.** Visualising A $\beta$  plaque distribution from AP +0.4 to AP -6.6 in a 12-month female 5xFAD mouse following Methoxy-X04 IP injection. Projecting Methoxy-X04 signal (magenta) over background autofluorescence (grey) reveals the lowest amyloid pathology in the cerebellum with most severe pathology in the SUB, cortex L5 and L6, and HPF. High Methoxy-X04 fluorescence over background signal combined with multichannel acquisition opens the door for whole-brain amyloid pathology tracking in Alzheimer's mouse models. **B.** Depth-coded GFP expression mapping from AP -4.0 to AP -2.0 in the Sox14gfp/+ transgenic mouse (P21) shows the previously unknown thalamic distribution in both first and higher-order sensory nuclei. Expression in the LGd fits with previously reported results. High-resolution imaging permits segmentation of single cells which can facilitate additional downstream analysis methods

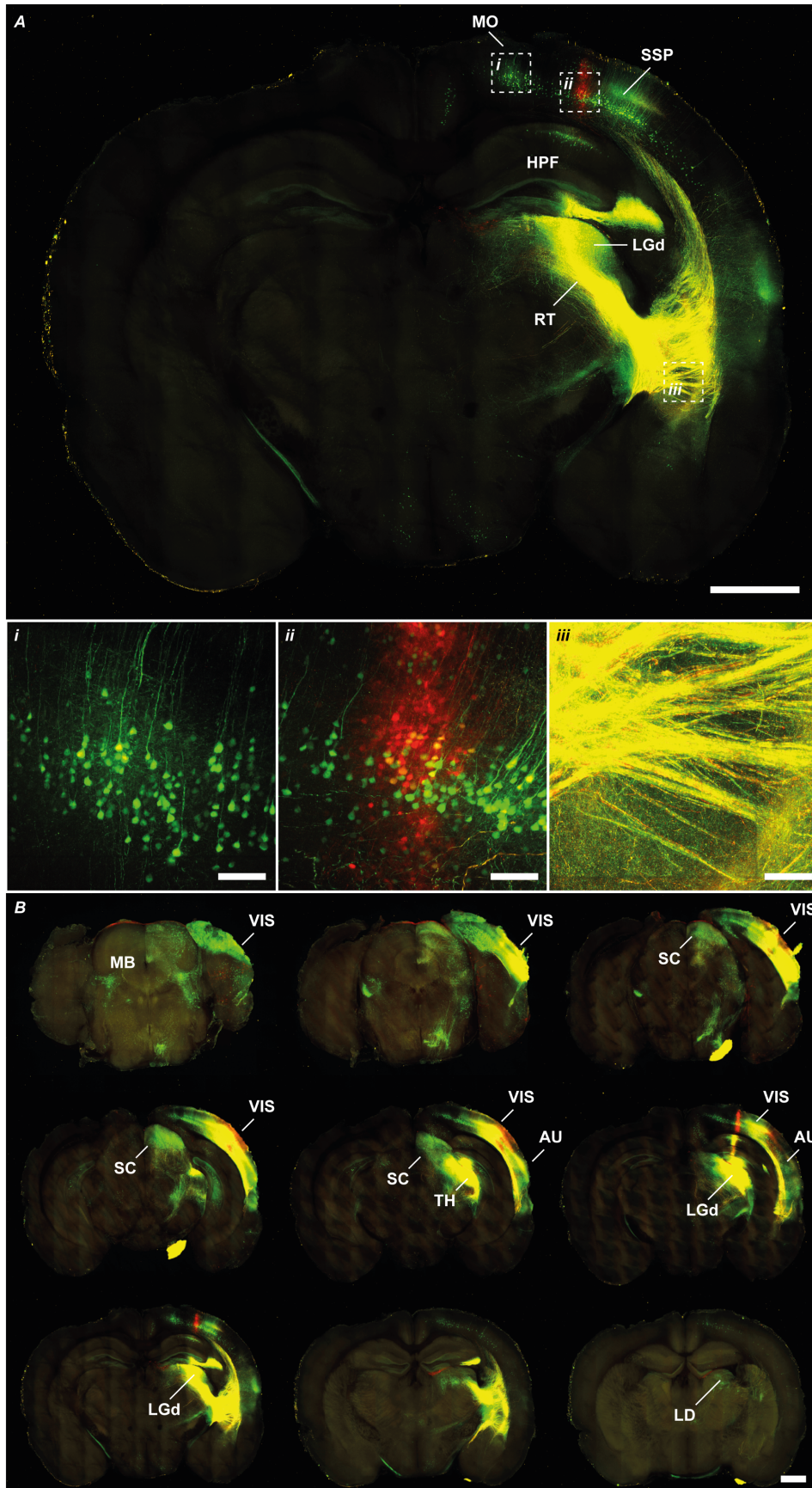


*such as cell counting. Subiculum, SUB; hippocampus formation, HPF; dorsal lateral geniculate nucleus, LGd.*

### **Co-injection of anterograde and retrograde virus labelling shows brain-wide connectivity of the LGd**

Co-injection of pAAV-CAG-tdTomato (anterograde) and pAAV-hSyn-EGFP (retrograde) in the LGd showed extensive mapping of axonal projections and soma throughout the mouse brain (Fig. 2.4). Labelled projections were located ipsilateral to the injection site (Fig. 2.4A) with some contralateral afferents projecting from the posterior midbrain (MB; Fig. 2.4B). Strong axonal and dendritic GFP expression (Fig. 2.4Ai) allowed morphology of individual neuronal processes to be studied. Multi-fluorophore tdTomato and GFP excitation was possible by choosing a 950 nm excitation wavelength, which balanced the cross-section absorption spectra for both fluorophores and gave strong GFP and tdTomato emission (Fig. 2.4Aii). The injection tract could be clearly identified by strong tdTomato expression in the somatosensory cortex (SSP), which is dorsally positioned to the LGd target injection site. Retrograde (green) and anterograde (red) labelling of projections from soma within the LGd show fascicle structuring into fibre bundle tracts (yellow) (Fig. 2.4Aiii). LGd afferent (green) projections mainly traverse through the reticular nucleus (RT). Pyramidal neurons in the visual cortex (VIS) layer 6 appear to be the primary afferent input to the LGd.

A mosaic of maximum intensity projections (Fig. 2.4B) shows distinct connectivity across an extensive anterior-posterior volume of the mouse brain. The ability to map this large volume demonstrates the utility of whole mouse brain imaging for the interrogation of brain-wide circuitry.



**Figure 2.4.** *pAAV-CAG-tdTomato (anterograde) and pAAV-hSyn-EGFP (retrograde) co-injection labelling of LGd connectivity. A. Maximum intensity projection of coronal volume of*

tissue from AP -2.3 to AP -1.8. Strong fluorescence signal from axons allows axonal morphology to be visualised (i). Multichannel STP imaging enables both anterograde and retrograde labelling to be distinguished (ii) with co-localisation of projections as they traverse through fibre tracts (iii). Panel A scale bar is 1 mm and 100  $\mu$ m for panels i-iii. **B.** Mosaic of maximum intensity projections across the mouse brain, revealing the extent of brain-wide labelling. Afferent and efferent projections mostly originate ipsilateral to the injection site with some MB afferent projections originating from the contralateral side. Scale bar 1 mm. Somatomotor cortex, MO; somatosensory cortex, SSP; hippocampus formation, HPF; dorsal lateral geniculate nucleus, LGd; reticular nucleus, RT; midbrain, MB; visual cortex, VIS; superior colliculus, SC; thalamus, TH; auditory cortex, AU; lateral dorsal nucleus, LD.

## 2.4 DISCUSSION

STP tomography provided an exceptional tool to study cytoarchitecture across the entire brain as well as specific volumes of tissue. Using a block-face imaging approach negates the necessity to manually align individual coronal sections of tissue, which improved the system throughput and easily allowed colour-coded intensity projections to reveal depth information in a two-dimensional representation. The custom stitching pipeline executed alongside STP acquisition generated a complete digital data set ready for analysis 10 minutes after acquisition completion. This expedited pre-processing timeframe was a considerable improvement over the original stitching pipeline. Cooling the PB and free-floating slices provided a general improvement to sectioning, imaging and slice integrity throughout the imaging experiments. Free-floating sections of tissue could be extracted and stored at 4°C for staining with DAPI, for example (Tarnowski et al., 1991).

Imaging at  $0.54 \times 0.54 \times 10 \mu\text{m}$  revealed fluorescence from individual soma, axonal and dendritic structures (Fig. 2.3B), and non-neuronal structures such as A $\beta$  plaque pathology (Fig. 2.3A). This proof of imaging capability could expand into several exciting studies such as A $\beta$  pathology quantification in Alzheimer's mouse models and comparison of the 5xFAD model with the less aggressive APP23 model. Cross-sectional studies on cohorts of Alzheimer's mouse models at a variety of ages would allow tracking of initial A $\beta$  seeding and spread across murine age. Furthermore, clear axonal and dendritic morphology from neuronal tracing techniques, can further our understanding of brain connectomics, cell types, and morphology.

STP tomography can therefore be used for several anatomical studies (see 'Chapter 5' and 'Chapter 6'). However, effort needs to go towards obtaining quantitative information from these extensive data sets. To this end, the following two chapters focus on the development of analysis tools to automatically register data with a reference atlas for identifying and delineating thousands of anatomical regions of the brain. Then we discuss the development of a novel pipeline which uses the registration result to target automated cell detection using a deep-learning approach, building on previous research (Moore et al., 2018). At the end of this thesis, these pipelines address the anatomical connectivity of local thalamic interneurons in the visual processing pathway, expanding upon the distribution observed in the Sox14gfp/+

experiment presented here, and for investigating anatomical changes in mice during the ageing paradigm.

## 2.5 TROUBLESHOOTING, CAVEATS AND FUTURE PERSPECTIVES

STP with a resolution of  $0.54 \times 0.54 \times 10 \mu\text{m}$  across the mouse brain took five days to complete and generated terabytes of data. Unfortunately, flaws in the system design and software implementation can result in several issues occurring during acquisition. Here we discuss some of these issues and any future considerations or resolutions that address those shortcomings, where applicable.

During STP imaging, the system saves several 3 MB tile images every second. Limitations in network bandwidth and inability to successfully save data can result in tile corruption. As part of the stitching process, a validation check is done on the tile data to ensure they can be opened without causing errors in the stitching process. If found, any corrupted tiles were replaced with a blank tile. Unfortunately, there is no way to reclaim the missing tile data and issues with reading and writing data can become more prominent for high Z-resolution scans as the number of successive optical images acquired per tile position and per second, increases. For a  $10 \mu\text{m}$  Z-resolution over a  $50 \mu\text{m}$  thick slice, five optical plane images were acquired in quick succession and resulted in a handful of corrupted tiles across the entire data set. Increasing the Z-resolution to  $5 \mu\text{m}$  over a  $50 \mu\text{m}$  thick, ten optical planes were acquired in quick succession resulting in eventual corruption of all tile data midway through the scan. The system is incapable of detecting corrupted images, meaning the scan will continue to operate without generating any image data. Network bandwidth is therefore essential to enable high Z-resolution STP scans.

The software was also prone to memory errors which caused the system to halt acquisition. Unfortunately, there is no easy way to restart the scan, instead requiring a new scan configuration. That said, it is possible that no data will be lost, as imaging can be restarted from the same slice position. This did require intimate knowledge of the system operation and which imaging parameters require resetting. Frustratingly, saving scan configurations does not save all necessary information meaning all parameters need to be reconfigured as if from scratch. In addition, stage translation position errors mean that a new scan can be slightly misalignment from the previous scan. Each new scan also requires a new output directory so data sets can be split across multiple directories. This can become quite challenging when handling several thousand image files.

A highly valuable technique in optical microscopy is image averaging. Line or frame averaging helps preserve signal of interest, whilst suppressing random noise fluctuations. Averaging is essential for improving signal-to-noise (SNR) level and therefore image quality. Unfortunately, the operating software for TissueCyte 1000 has no averaging capability, and it is not possible to force the system into an averaging scheme. Of course, averaging would increase imaging time by a factor of the number of averaged images.

The laser control and software control are completely independent, meaning only a single laser excitation wavelength is possible. Having the ability to switch laser excitation wavelength would enable peak excitation of multiple fluorophores, rather than rely on a single excitation wavelength which balances between the peak cross-sections of multiple fluorophores in the sample.

Control of the motorised XYZ stage during the alignment stage of system operation is done using individual distance commands which are input as a single numeric value. To translate the stage +1000  $\mu\text{m}$  in the X-axis and -2000  $\mu\text{m}$  in the Y-axis requires first inputting the value 1000 and moving the stage in the +X direction, and then changing the value to 2000 and moving the stage in the -Y direction. Whilst stage translation is fast, repetitive stage translation for alignment of the sample is time intensive. To further complicate the process, there is no implementation of error or collision detection. Care must, therefore, be taken to prevent accidental collision of the sample with the objective or microtome. Stage movements are required to be small if operator experience is limited with the system, adding to an already time-intensive procedure. A valuable implementation would be the addition of a control panel with three dials for three-dimensional manipulation of the stage to enable faster alignment protocols for the system.

Microscopes using PMT detection should be light-shielded due to their sensitivity. However, there is poor design of the TissueCyte 1000 system in this regard. The system is contained within an enclosure which is surrounded by a laser-blocking blackout curtain. Whilst this shields the system from external light during imaging or during standby, optimising the imaging parameters during set up requires frequent access to the system, which is inside the

enclosure, whilst also requiring the system to be light-shielded for the duration of the imaging. Crucially, the light path for the system is not light-shielded and more importantly, the PMTs are not shielded. Redesigning the system to a more optically compact layout with light shielding between optical components would enable imaging without the constant concern of damaging the PMTs.

A significant component of block-face imaging is the ability to section the tissue. Modifications were made to optimise slice sectioning quality from 100  $\mu\text{m}$  down to 40  $\mu\text{m}$ . However, the microtome vibration system is noisy, and slicing could be further improved to robust sectioning of 20-25  $\mu\text{m}$ . A modified microtome instrument would increase slicing quality and time efficiency. The use of a diamond or sapphire blade would also permit slicing down to even 10-15  $\mu\text{m}$ . This would also benefit slicing of tough tissue structures, such as the membrane meninges, which might otherwise add to imaging artefacts.

Despite fast manipulation, the motorised XYZ stage exhibits severe errors in the reported position. This is due to a driver issue which is still required by the operating software of TissueCyte 1000. Knowing the exact stage position can be used to improve the stitching pipeline down to single pixel accuracy and ensures complete reconstruction of images. This could be resolved by completely swapping out the motors for the stage control with newer and faster components and utilise the correct error readout to improve the stitching pipeline.

Obtaining a complete coronal section of tissue requires tiling the objective FOV over a defined tile area (Fig. 2.2C). However, the shape of a coronal section of tissue does not fill the complete tiled area, meaning there is redundant imaging. Considering the oblate shape of a mouse brain, a considerable number of blank tiles stacks up over the course of volume imaging. STP imaging time could be expedited with a quick first pass scan at low resolution to obtain the exact shape of the tissue section that requires imaging. High-resolution imaging could be targeted to tile positions where there is tissue, with blank tiles omitted.

Ultimately, a lot of the problems with the system as described above originate from issues with the operating software. A recommended area of future direction is, therefore, the complete overhaul of the operating software and a redesign of the system to improve several



mechanical drawbacks. Creating a more compact microscope would improve usage protocols, reduce risk of damaging the PMTs, improve the XYZ stage control, as well as imaging and position readout. Faster mechanical manipulators would also expedite the imaging process. Whilst this would be an extensive endeavour, the STP system created by TissueVision is fundamentally limited in many areas and could be drastically improved for speed, accuracy and image quality.

## **WHOLE-BRAIN MOUSE REGISTRATION TO COMMON COORDINATE FRAMEWORK FOR ANATOMICAL DELINEATION**

### **ABSTRACT**

One challenge associated with whole mouse brain digital data sets is the delineation and extraction of anatomical regions of interest. Human delineation is subjective, error-prone, and challenging on large data sets containing thousands of individual images. Segmenting a three-dimensional data set is also further complicated when using a two-dimensional reference atlas or image. To this end, we created an automated registration pipeline using the Elastix framework, which registers high-resolution serial two-photon tomography data of the whole mouse brain with the comprehensive Common Coordinate Framework version 3 (CCF3) atlas from the Allen Brain Institute. Executing the pipeline on downsampled (10  $\mu\text{m}$  isotropic) imaging data takes 30 minutes and results in the identification of thousands of anatomical regions referenced according to the Allen Brain Institute taxonomy. As registration procedures derive from MRI applications, we optimised parameters to ensure robust performance on high-resolution data with comparisons to existing parameter configurations used in other studies. The custom parameter configuration created here presented the most optimum registration procedure, showing excellent performance in target registration error, image similarity measures and visual assessment. Improved registration was also observed on tissue presenting with tearing and corruption.

### 3.1 INTRODUCTION

Whole-brain imaging techniques including STP tomography, light-sheet tomography, MRI and PET, come with a variety of analytical challenges. One of the most critical of these is the anatomical delineation of the tissue subject, as it permits inference on structural and functional topography (Ding et al., 2016; Nakamura et al., 2000). It is also necessary for providing a common spatial framework to compare subjects within a species, or even comparing between species. The common, classical approach for tissue delineation is using expert human anatomists to annotate regions of interest in tissue, visually. Staining methods can assist this challenge by adding contrast between neuronal structures, a technique perpetrated by the likes of Ramón y Cajal when investigating the laminae tonotopic organisation of the auditory cortex (Ramón y Cajal, 1909). Staining in the modern laboratory still relies on these classic protocols such as the Nissl stain technique (Nissl, 1894).

Use of Nissl and other staining protocols such as Acetylcholinesterase staining (Karnovsky & Roots, 1964), in combination with expert human neuroanatomist annotation, lead to the creation of the Franklin and Paxinos (FP) atlas (Franklin & Paxinos, 2008) and also the Allen Reference Atlas (ARA; Dong, 2008), the first generation of atlas from the Allen Brain Institute. Both are arguably the two most commonly utilised mouse brain atlases. The FP atlas consists of 100 coronal sections evenly spaced 120  $\mu\text{m}$  apart, spanning the complete rostral-caudal extent of the mouse brain, with over 800 anatomical delineations identified. In comparison, the ARA consists of 132 coronal sections evenly spaced at 100  $\mu\text{m}$  apart and also spans the full rostral-caudal extent of the mouse brain. The ARA also has a sagittal atlas consisting of 21 slices spaced at 200  $\mu\text{m}$  apart, spanning the medial-lateral extent of the brain.

The FP atlas is used ubiquitously for stereotaxic injections and electrode placement. Consequently, the FP atlas can be regarded as the most universally used coordinate system and enables different research laboratories to share, compare and standardise protocols for stereotaxic surgeries and injections. That said, biological variation with overall brain size and differences in subregion volumes means that a simple coordinate position is not directly translatable between different subjects. Stereotaxic injection techniques take several months to practice with multiple calibration steps required to ensure the target brain region is targeted. On the other hand, the ARA atlas is frequently used for reference in the

identification of tissue regions during slice work. This is due to the ARA atlas constituting one of the most publicly accessible atlases from those that are available. Whilst the ARA is comprehensive for its annotation, the atlas itself is not readily associated with a common coordinate scheme much like the FP atlas.

In 2015, the Allen Brain Institute released a new atlas which used high-resolution STP tomography consisting of 528 coronal sections spaced 25  $\mu\text{m}$  apart, of 1,231 samples of the C57BL/6J mouse (Kuan et al., 2015). All samples were registered together to generate a single average atlas, which was referenced with the ARA to generate a segmented annotation atlas. The 528 coronal sections were interpolated in the rostral-caudal axis to generate 1,320 coronal sections for a complete annotation atlas of 10  $\mu\text{m}$  isotropic voxel size. The atlas was intentionally created to provide a common coordinate framework (CCF) and presented the most comprehensive atlas at the time with a virtually complete mouse brain sampled every 10  $\mu\text{m}$  in all axes, instead of being restricted to sampling every 100  $\mu\text{m}$  or 120  $\mu\text{m}$  as in the original ARA and the FP atlases. The lack of sampling in the rostral-caudal domain for the ARA and FP atlases is particularly problematic when working with slice tissue which does not spatially correspond with any of the atlas coronal sections.

Additionally, the developing mouse brain has structural differences with an adult mouse (Baloch et al., 2008) and general biological variability between samples, including brain size, ventricle sizes, and slight deviations in the position and shapes of many brain structures. Such variations can make it challenging to use the lower-sampled FP and ARA atlases. Adding to this, tissue samples derived from subjects with a neurological disease pathology may exhibit vastly different brain structure and volume, as seen in hallmark subjects of Alzheimer's disease (Maheswaran et al., 2009) and Huntington's disease (Cepeda-Prado et al., 2012). Extraction of brain tissue during transcatheter perfusion techniques can also deform, tear, and corrupt tissue integrity. Having access to a higher rostral-caudal sampled atlas gives additional contextual anatomical information which can aid with manual referencing of tissue regions.

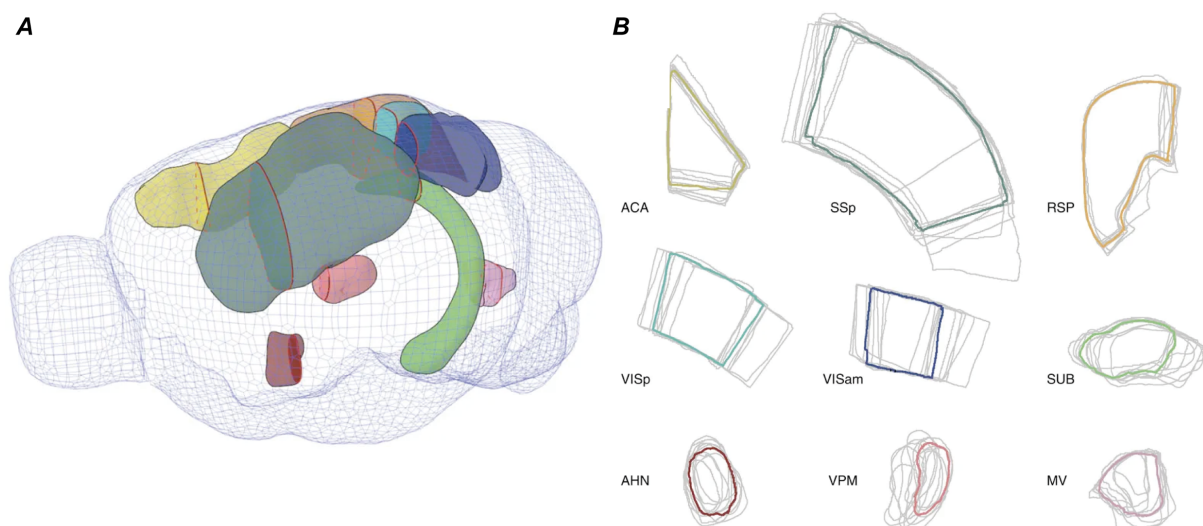
A significant use of brain-wide atlases is to compare samples by providing a common spatial framework. The comprehensive CCF atlas excels in this regard. However, the CCF atlas annotation labelling is derived from the FP atlas whilst the annotations boundaries are

derived from the ARA. Consequently, the FP and ARA labelling and annotations do not conform with each other. The most recent 2017 CCF atlas from Allen Brain Institute (CCF3) improved upon the 2015 version by increasing the interpolation accuracy between Z-slices and, perhaps prudently, also changed the labelling scheme from the ARA, from which it was initially derived (Q. Wang et al., 2020). There is, therefore, a remarkable discrepancy between all existing mouse atlases. Yongsoo Kim Lab (YKL) attempted to address this inconsistency by merging the annotations and labels in the CCF3 and FP atlas (viewable online at <https://kimlab.io/brain-map/atlas>) (Chon et al., 2019). Atlas region coverage was also improved by also including the cortico-striatal projection pattern segmentations in the dorsal striatum. Although the atlas achieves the goal of merging atlases, it does not contain the extensive 1,320 coronal sections of the CCF3 atlas and is limited to only 132 Z-slices. This is comparable to the ARA and FP atlases. Consequentially, the CCF3 atlas is still better equipped for whole-brain studies. Despite recent efforts in atlas generation, a common unified atlas does not currently exist.

Referencing an atlas for tissue delineation adds another complication. That is expert human neuroanatomists present annotation discrepancies between themselves and each other (Niedworok et al., 2016). An anatomist might use high contrast landmark features in tissue to assist in the demarcation of tissue regions, such as white matter tracts and ventricles. However, protocols detailing how to identify these anatomical landmarks and how to delineate them are variable (Nestor et al., 2013).

One study tasked 11 expert human anatomists with annotating nine regions in coronal sections of mouse brain tissue (Fig. 3.1; Niedworok et al., 2016). Regions included the anterior cingulate area (ACA), anterior hypothalamic nucleus (AHN), medial vestibular nucleus (MV), retrosplenial cortex (RSP), primary somatosensory area (SSp), subiculum (SUB), primary visual cortex (VISp), secondary visual cortex anteromedial part (VISam) and ventral posteromedial nucleus of the thalamus (VPM). They observed weak agreement between the human anatomists for every tested structure with worst performance seen in the VPM where some annotations barely shared any overlap with the other anatomists. The ACA, RSP, SSp, VISp and VISam were located in cortical brain areas meaning delineation was assisted by a sharp contrast between the tissue and the background signal at the surface of the neocortex.

However, those structures show variability in human delineation once the boundary was no longer traced at the surface (Fig. 3.1B). The remaining regions reside in the brainstem (MV), hippocampus (SUB), thalamus (VPM) and hypothalamus (AHN) of which the SUB provided the least challenging annotation task, most likely due to an abundance of contrasting landmarks in the hippocampal formation. Overall, the AHN, VPM and MV, presented the most challenging regions. Although testing was performed with only 9 regions and 11 human anatomists, there is an apparent significant problem arising from inter-human variability when it comes to delineation of brain regions. It should also be noted that the term “expert” was used to cover performers from different backgrounds and expertise in anatomy. Sometimes, it may not be practical to rely on expert anatomists to perform delineation, and instead rely on someone less skilled.



**Figure 3.1.** Human manual segmentation error for nine anatomical test regions in the mouse brain. **A.** 3D illustration of nine structures in the mouse brain identified in the Allen Common Coordinate Framework atlas. **B.** Segmentation traces by 11 human raters (grey) with the averaged outline (bold) for the following structures: anterior cingulate area (ACA); primary somatosensory area (SSp); retrosplenial cortex (RSP); primary visual cortex (VISp); secondary visual cortex, anteromedial part (VISam); subiculum (SUB); anterior hypothalamic nucleus (AHN); ventral posteromedial nucleus of the thalamus (VPM); medial vestibular nucleus (MV). Figure adapted from Niedworok et al., 2016.

The first analysis hurdle when working with high-resolution STP data sets is the reliable and accurate delineation of structures of interest. Manual labelling is evidently challenging and subject to human variability. So, how about an automated registration approach? Many state-of-the-art registration toolkits exist, including Automated Image Registration (AIR; Woods et al., 1992), Advanced Normalization Tools (ANTs; Avants et al., 2014), Elastix (Klein et al., 2009), and NiftyReg (Modat et al., 2014). These toolkits were developed to solve the challenge of biological image data registration.

However, the prevalence of MRI techniques means these toolkits were designed mostly with a focus on MRI data. They were also created in a time which pre-dated the advent of high-resolution data sets such as those created with STP tomography. An MRI data set for a mouse brain in a volume of  $120 \times 100 \times 200$  mm can have an acquisition voxel size of  $1 \times 1 \times 1$  mm which equates to approximately 2.5 million total voxels and yields around 4 MB of storage. The high-resolution STP data used in this study has an approximate  $0.5 \times 0.5 \times 10$   $\mu\text{m}$  voxel size, which equates to 960 trillion voxels for the same tissue volume and takes up 3 TB of storage. Execution of existing registration toolkits on a data set of this magnitude is impossible. A minimum requirement of 3 TB of RAM is necessary just to load the data; a powerful modern research workstation may only have a respectable 128 GB RAM. This is also ignoring the RAM requirement for registration procedure itself. A more realistic approach for STP data registration is to use data in a similar format as the Allen CCF3 atlas with an isotropic 10  $\mu\text{m}$  voxel size. This data set equates to 2.4 quadrillion voxels for the same volume as the MRI data set. As the dimensional scaling between MRI data and STP data is so vast, diligence is needed to ensure registration methods based on MRI data are robust for higher-resolution data sets. A comparative study of commonly used intensity-based registration methods revealed ANTs and Elastix to be the best performers when using cleared mouse tissue deformed with a Gaussian vector field (Nazib et al., 2018). Elastix also required less computational performance.

With this in mind, we created an Elastix based registration pipeline written in Python and using the SimpleElastix framework. The pipeline goal was the registration of whole mouse-brain STP data with the Allen CCF3 atlas space using an average brain atlas and a corresponding annotation atlas. Following the calculation of the transformation map

calculated for registering a downsampled STP data set with the average CCF3 atlas, the transformation was subsequently applied to the annotation atlas. This resulted in the segmentation of the STP data into thousands of anatomical regions as labelled according to the Allen Institute structure taxonomy. In addition, we created a hemisphere atlas to distinguish nuclei between the hemispheres. This was otherwise not possible with the CCF3 annotation atlas alone. To evaluate the optimal parameter configuration for robust registration results, we tested several parameter configurations, including a new parameter configuration explicitly created for this challenge. Registration errors were compared to a human target error performance and visually assessed for a more accurate qualitative performance. The Elastix registration principles are outlined below.

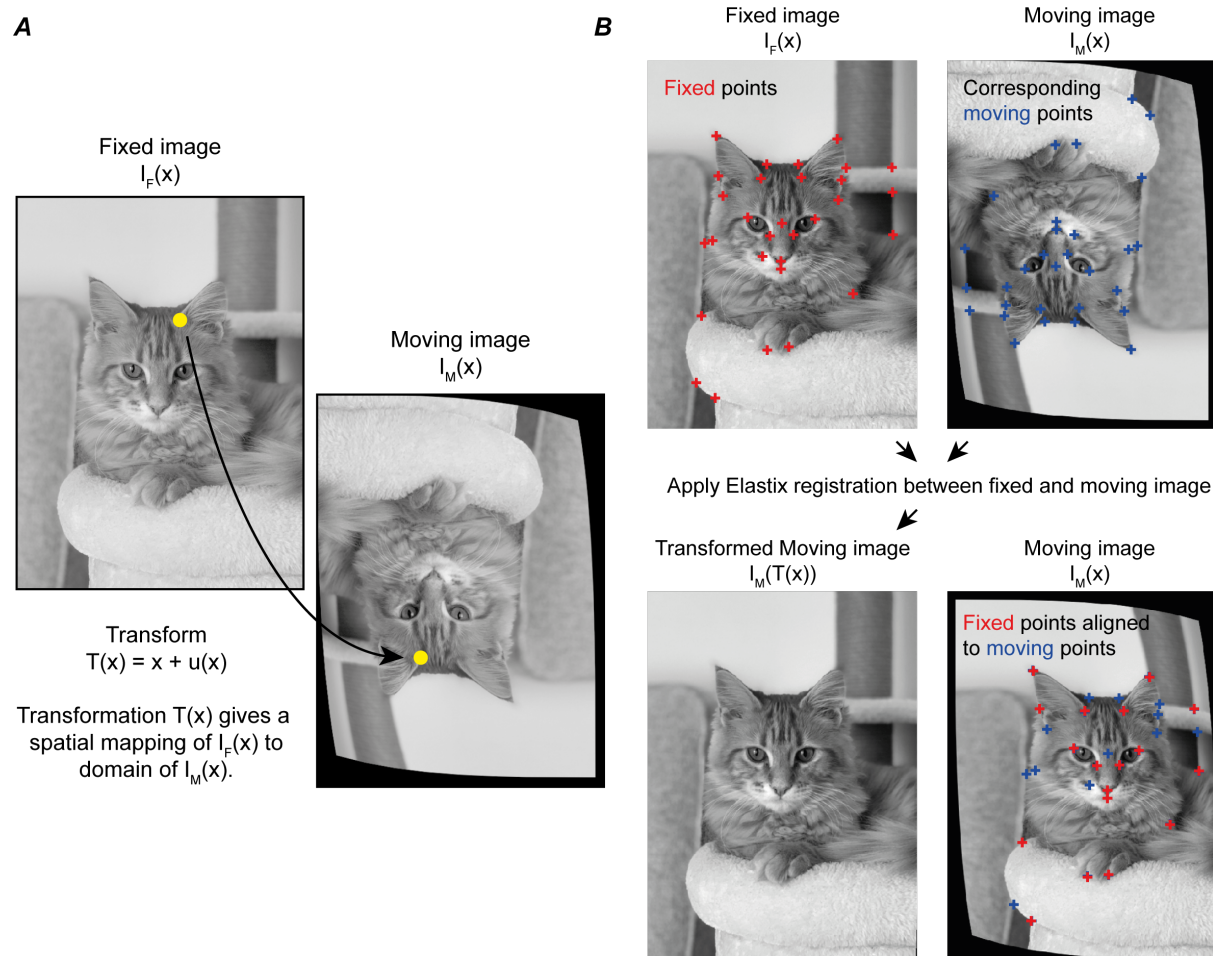
### **Elastix Registration**

Elastix is a medical image registration toolkit which allows a moving image to be spatially normalised to a reference image (Klein et al., 2009). Elastix, and intensity-based registration in general, relies on a set of sequential transformation steps of increasing displacement parametric complexity to iteratively displace the moving image to match a reference image. For Elastix, these steps are rigid, affine and non-rigid transformation. Rigid transformation comprises rotation and translation displacements. Affine transformation comprises rotation, translation, scaling and shearing displacements. A non-rigid transform deploys a deformable grid over the data to elastically manipulate a set of local control points using a k-order spline function for interpolation of pixel values. Non-rigid is only required if global alignment is suitable for the challenge at hand, and finetuning in local alignment is not needed.

Registration is a nuanced research area which sometimes uses the same descriptors for referencing different specifics. To be more understandable, we acknowledge some of these misconceptions in the following explanation of the registration process. The standard registration procedure involves a fixed image  $I_F(x)$  and a moving image  $I_M(x)$ . The fixed image can often be referred to as a reference image as it does not exhibit any deformation during the registration procedure. This can cause confusion as a reference may also refer to a template or atlas which is referenced for the labels after registration is complete. A transformation,  $T(x)$ , is calculated using a sequential combination of rigid, affine and non-rigid



transformation displacements,  $u(x)$ , to generate correspondence between pixels in the fixed image with those in the moving image (Fig. 3.2A).



**Figure 3.2.** Principles of registration. **A.** Registration between a fixed,  $I_F(x)$ , and moving,  $I_M(x)$ , image. A transformation,  $T(x)$ , is calculated by displacing,  $u(x)$ , the moving image such that a point in the fixed image can be spatially mapped to a point in the moving image. Yellow markers indicate the spatial correspondence between the fixed and moving image. **B.** Transformations can be applied to a set of coordinate points for evaluating target registration error (TRE) as a Euclidean distance between corresponding landmarks. Sparse points transformation requires a landmarked fixed image  $I_F(x)$  (top-left; red markers) and a moving image  $I_M(x)$  with corresponding landmarks (top-right; blue markers). During Elastix registration, the moving image is registered with the fixed image  $I_M(T(x))$  (bottom-left). The transformation  $T(x)$  is applied to the fixed points in  $I_F(x)$  to map them to the moving image domain  $I_M(x)$  (bottom-right; red and blue markers show alignment). Initial TRE  $39.61 \pm 9.74$  (Mean  $\pm$  SD,  $n=30$  landmarks) pixels with final TRE  $0.96 \pm 0.60$  pixels. Note that the

*transformed markers align within a mean error of a single-pixel width which makes it impossible to distinguish between both sets of markers (red and blue) in this representative example.*

Each Elastix transformation step looks to optimise a similarity measure to assess how well the registration is performing. This metric quantitatively assess a measure of correspondence between the two images as the deformation is taking place (described in more detail under 'Registration Components'). If the correspondence is weak, either because the registration parameter choice is inadequate, or because the registration is in early stages of execution, then the moving image is iteratively deformed according to the rigid, affine or non-rigid transformation. Deformations repeat until new similarity measures yield improved scores. This iterative improvement to registration is driven by a gradient descent algorithm (described in more detail under 'Registration Components') which mathematically uses a similarity measure formulated as a cost or optimisation function. Ultimately, the goal of registration is to minimise the cost between the fixed and moving images by optimising the final transformation to the moving image.

Registration is a mathematically driven process, meaning there is no understanding of image context. Large deformations in a moving biological image are likely to be inaccurate but could still lead to an improved similarity measure. With Elastix, it is possible to use a penalisation term that restricts the size of deformations and prevents inaccurate deformations. Additionally, modifying the grid control point spacing allows control of the precision of non-rigid registration. Smaller grid spacing is useful for aligning smaller structures at the cost of computational complexity and execution time. A grid spacing which is too-fine can generate wave-like distortions in the moving image as the algorithm is encouraged to match as many locally distributed pixels as possible.

The annotation CCF3 atlas is a 32-bit data set and comprises of a unique pixel intensity per anatomical structure. As this has no pixel correspondence with the autofluorescence pixel intensities in an STP data set, it is not feasible to perform the registration using the annotation atlas. Therefore, registration is performed using the average CCF3 atlas (16-bit) with the resulting transformation applied to the annotation atlas. An essential nuanced understanding

of the registration processes is that the moving image is the one which is deformed to align with the fixed image spatially. However, the transform is calculated as a mapping from the fixed image domain to the moving image domain. This is a significant appreciation of how the registration process works. Even though the CCF3 atlases provide the common spatial framework, registration is intentionally applied with the CCF atlases constituting the moving image and the STP data set constituting the fixed image. The final transformation map generated by the registration process gives a mapping of a location in the fixed image to a location in the moving image. It is thereby allowing the STP data to become spatially mapped to the CCF3 atlas data. The registration operation is defined in this manner because the moving image becomes displaced and can either lose pixel positions or have pixel positions generated through interpolation. Dissemination or interpolation of pixels means the moving image loses its spatial correspondence. We can, therefore, map every pixel in the fixed image to a position in the moving image because we have not lost any pixel correspondence in the fixed image which remains undeformed. If we tried to map a pixel from the moving image to the fixed image, it is possible the pixel to be re-mapped gets lost during the transformation and does not spatially correspond with any position in the fixed image.

Elastix can also use corresponding points between fixed and moving images to either assist transformation or to be transformed using the final transformation map resulting from registration. In the first case, corresponding points between the fixed and moving image can be used to encourage Elastix to minimise the Euclidean distance between the corresponding points. This approach for registration is useful if a set of known points can be routinely identified between the images. In reality, this requires manual work to identify regions, and for large 3D data sets, this is time-intensive and introduces human errors in identifying landmarks. Additionally, STP imaging data of brain samples are rarely perfectly aligned during imaging, meaning nearby landmarks in the fixed image set may be distributed widely across the STP data.

The second use of corresponding points enables a transformation to be applied to points in the fixed image to identify their spatial correspondence in the moving image domain. Whilst the transformation deforms the moving image to match the fixed image, the same transformation result maps from the fixed image to the moving image domain (Fig. 3.2B).

Target registration error (TRE) can, therefore, be calculated using corresponding points. For corresponding points,  $a$  and  $b$ , where  $a = (a_1, a_2, a_3)$  and  $b = (b_1, b_2, b_3)$  in a 3D cartesian space, the Euclidean distance  $d(a, b)$  is calculated with the following.

$$d(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (\text{Eq. 3.1})$$

The TRE is expressed as a mean of the Euclidean distances,  $d(a, b)$ . Evaluating TRE using corresponding points is limited by the number of points and is highly subjective, given the structural complexity of the data to be registered. Furthermore, the location of landmarks is also crucial for achieving a reflective TRE. Landmarks chosen on high contrast boundaries between structures (the boundary of skull and background in MRI data, for example) will be easily aligned during registration. However, they will not give any indication for success in internal brain tissue alignment (grey matter, white matter and substructures, for example), without those lower contrast regions also requiring landmarking.

A detailed description of Elastix parameters is in the Elastix manual (available online at <https://github.com/SuperElastix/elastix/wiki>). Despite the complexity of image registration, studies performing registration with Elastix are usually sparing on parameter details. Trying to emulate the success of registration is therefore very challenging, and often there is no intuition as to why parameters are chosen to be as they are. That said, a select few papers that have used Elastix in their research, have published their parameter configurations online ([http://elastix.bigr.nl/wiki/index.php/Parameter\\_file\\_database](http://elastix.bigr.nl/wiki/index.php/Parameter_file_database)). However, these parameters follow a very similar template suggesting they are derived from the same example file or a previous study. In an attempt to be more open about parameter choices, the key components and their effect on registration, are detailed below.

### **Multi-resolution approach**

Fixed and moving images might require alignment across a multitude of scales. For example, the isocortex of a mouse brain can be considered a large global structure as it spans a large volume in the medial-lateral, anterior-posterior and dorsal-ventral axes. On the other hand, subcortical structures, such as fibre tracts which are only a few microns wide in cross-section, might be considered part of the finer, local scale. Accurate registration across all scales is not

guaranteed using a single scaled data set. Hence, we require a multi-resolution approach. Multi-resolution with Elastix is applied by using a smoothing image pyramid which applies a Gaussian kernel to the image at different scales, representative of downsampling. Alternative methods include a recursive image pyramid which applies both smoothing and downsampling, and a shrinking image pyramid which applies the downsampling step, only. This component of the registration is utilised by dictating the number of resolutions and the smoothing (downsampling) amount at each resolution step. As an example, a chosen number of resolutions of 4 for a 2D image will by default have a smoothing schedule of (8 8 4 4 2 2 1 1) meaning for the first resolution level the images are smoothed with a Gaussian kernel of  $\sigma=8/2$ , then  $\sigma=4/2$  at the second level,  $\sigma=2/2$  at the third level, and finally  $\sigma=1/2$  in the last resolution, where the last resolution is also the original scale of the image. As the image is 2D there is equal scaling of the Gaussian kernel across the x and y dimensions, hence the pairs (8 8) and (4 4) and so on. In 3D the schedule would become (8 8 8 4 4 4 2 2 2 1 1 1). Any scale factor can be chosen and does not have to be equal for all dimensions at any level.

Using multiple resolutions for registration is critical to achieving accurate results across all scales. Using two resolutions can result in much better results whilst using four, five or six resolutions do not appear to affect the result substantially between themselves. Adding resolutions also increases execution time. If only working with global alignment, fewer resolutions are required. For fine-structure alignment, higher resolutions of four and more are ideal.

## **Transformations**

An initial rigid registration step is performed using the Euler transformation, which accounts for global rotation and translation displacement. The Euler transformation is followed by an affine transformation which accounts for rotation, translation, scaling and shearing displacement. Lastly, a non-rigid step is performed using a k-order b-spline displacement. Of the three stages, the non-rigid b-spline transformation is the most complex. Here a free-form deformation field is deployed over the moving image allowing it to be elastically manipulated using a set of control points. Grid control spacing is a critical component at this stage. The finer grid sizes allow for finer tuning of the transformation at the expense of computational complexity and time. Finer control does not necessarily mean higher registration accuracy as

biologically inaccurate deformations could generate better similarity scores. If combined with the multi-resolution strategy, the grid points can also be scales according to the resolution by specifying a schedule similar to the smoothing schedule. Different grid spacing can be dictated at different downsampling scales to modify the fidelity of control.

The transformations stages should be executed in the order of rigid, affine and non-rigid (if non-rigid is required). Although affine transformation also handles rotation and scaling displacement performed in the rigid Euler transform, we observed failure to align the moving data globally when using affine and non-rigid transformation only. Including rigid Euler transform as a prior for the affine transformation resulted in robust registration results. An initial rigid transformation is recommended to avoid getting the registration stuck in a weak initial alignment. This is particularly important when using data sets which require substantial gross alignment, as is the case with STP whole mouse-brain data.

It is also possible to force an initial alignment by using a centre of gravity alignment. This procedure is effective when working with mouse brain data which may not cover an extensive rostral-caudal volume.

### **Similarity measures**

The similarity measure is an essential component in the registration procedure and significantly impacts the result. The metric samples intensity values in randomly chosen regions of the fixed and moving image and calculates a measure of correspondence between the two. If the correspondence is high, then registration is going well. Likewise, if they are low, the transformation needs to be directed to improve. Only two metrics are considered in the majority of registration applications; Advanced Normalised Correlation (ANC) and Advanced Mattes Mutual Information (AMMI). Both of these functions are reworked into an objective or cost function such that high similarity between the images results in a low cost. The cost drives the transformation in the direction of improved accuracy.

ANC (also termed Normalised Correlation Coefficient) is a cross-correlation calculated between pairwise pixel intensities. The score is normalised by the square root of the

combined cross-correlation of the fixed image and itself, and the transformed moving image and itself.

AMMI calculates the mutual information given by the pixel intensity between the fixed and transformed moving image and represents how well the transformed moving image intensities predicts the fixed image intensities. As the calculation uses a probability density distribution, the metric is suitable in cases where the fixed and moving image intensities are quite varied, possibly due to different imaging modalities, for example.

### **Optimiser**

An optimiser seeks to minimise a cost or objective function which is described by the similarity measure. The most commonly used optimisation function is gradient descent and is used in a multitude of optimisation tasks, including machine learning and registration.

Gradient descent is an iterative approximation of the differential of the cost function, which seeks to optimise parameters in the cost function which lead the gradient descent algorithm to converge on a minimum. At this point, there is an optimisation of parameters in the transformation, which leads to the best similarity score.

Gradient descent is a complex iterative procedure which is computationally intensive. The adaptive stochastic gradient descent algorithm circumvents some of the bottlenecks by initiating the registration with predefined parameters and uses an approximation of the total similarity cost by using sampled regions in both the fixed and transformed moving images.

In general, the adaptive stochastic gradient descent approach works well.

### **Interpolator**

As the moving image is iteratively warped, the pixels in the moving image require interpolation to fill in values with respect to the other control points. Both linear interpolation and b-spline interpolation perform well for registration of biological image data. Linear interpolation simply calculates the value for pixels as a weighted average of intensities in the surrounding area, where the weights of each local pixel are calculated from the distance to

the pixel being manipulated. B-spline interpolation uses a k-order spline function. A first-order spline interpolation function is simply a linear interpolation. In general, a third-order b-spline interpolates pixel values quite well as it compensates for local fluctuations in the intensity, which might be reflective of local anatomical structures that should not be lost or corrupted during registration.

Another interpolation method is the nearest neighbour interpolation. This approach simply duplicates the neighbouring pixel value and is not advised for biological data. However, for applying transformations to labelled annotation data, nearest neighbour interpolation prevents label corruption during the transformation as interpolated intensity values do not match the unique labelling of structures in the atlas.



## 3.2 METHODS

Here, we describe the methods for the preparation of data prior to registration, the registration procedure itself, including a comparison to previous studies' parameters, and the final result of the whole-brain registration pipeline. For registration and segmentation of thousands of anatomical structures in the STP data, we used the Allen CCF3 average and annotation atlases, given its more comprehensive rostral-caudal sampling.

### Choosing parameter maps

To dictate the configurations for registration, Elastix reads in a parameter map file with parameters listed as key-value pairs. A parameter map was required for each transformation; rigid, affine and non-rigid b-spline. Previous studies have used Elastix for whole-brain registration with data from MRI, of which the most commonly used comes from Hammelrath et al. (2016) who imaged the intact mouse brain inside the skull at a voxel resolution of  $0.14 \times 0.14 \times 0.40$  mm and dimensions of  $128 \times 128 \times 22$  pixels. The parameter configuration files, named Par0025, contain three parameter maps for rigid, affine and b-spline registration.

Two other parameter maps, Par0033 from Bink (2016), and Par0038 from Mulder et al. (2017) have also been used for whole mouse brain MRI registration. The Par0033 parameter files were used on MRI data with voxel size  $0.07 \times 0.07 \times 0.07$  mm, whilst the Par0038 configuration from Mulder et al., was used on multiple MRI data sets with voxel sizes in the range 0.10 to 0.12 mm in XY and 0.50 to 0.80 mm in Z. Each parameter file was considered for the registration process to assess which configuration would work best for the higher resolution STP data.

As each parameter file originated from MRI studies with voxel sizes much larger than the 10  $\mu$ m isotropic voxel size of the CCF3 atlases, all three parameter files were modified for their scaling. Firstly, the element spacing of the input data was set to 0.01. As the Elastix toolkit is derived from the ITK toolkit, which was initially developed for MRI data, all dimension default to units of millimetres unless explicitly defined. Therefore, the chosen element spacing value of 0.01 translated to a voxel size of 0.01 mm, which was equivalently 10  $\mu$ m. It should be noted that properties about the input data set such as element spacing can be saved as image metadata if the file format allows. These properties are automatically read during Elastix

instantiation, which negates the need to specify their values. However, we found a bug with ImageJ/Fiji when saving image properties as metadata. Information defined under the image properties in ImageJ/Fiji is correctly saved to metadata compatible files except for element spacing. This was true for TIFF, MHD and MHA file formats. Therefore, setting the element spacing parameter manually in the parameter map files was necessary to ensure the correct spacing was guaranteed.

The parameter maps also defined the grid spacing for the non-rigid transformation in terms of physical units of mm. Given the previously noted issue with element spacing and to prevent runtime issues that might arise from a mismatch in dimensions and units, the grid spacing was changed to be defined in voxel sizes for all parameter maps. In the case where the grid spacing was defined as voxels, the physical unit size for the MRI data was calculated and then converted into a voxel spacing given the 10  $\mu\text{m}$  isotropic voxel size of the downsampled STP data. If the parameter map already defined the grid size as physical units, the conversion step was only necessary.

The next modification to the parameter maps ensured stability in the registration when STP data might exhibit gross global misalignment with the CCF3 atlas or when there is a different rostral-caudal volume between the STP and the CCF3 atlases. Two initial calibration parameters were added to each file; an automatic transform initialisation and the centre-of-gravity approach. A combination of these two parameters forces an initial global alignment based on the centre of mass of the fixed and moving data. Other initialisation methods include, by origin, and by the geometrical centre. However, as the origins are both defined as the most dorsal-caudal-lateral point in both the fixed and moving data sets, aligning by origin would not align the overall data representing the tissue. In addition, the geometrical centre would be considered the central point of both data sets, which again does not take into account the location of the mouse brains within the atlas space.

The final modification to each parameter file was a limitation to the number of Jacobian measurements, which is used to calculate a covariance matrix between the fixed and moving image during cost function minimisation. Typically, the number of measurements is automatically determined at runtime. However, for parameter maps Par0033 and Par0038,

the number of calculations when determining the Jacobian matrix caused stalling in computation. Therefore, the number of Jacobian measurements was set to 1,000 for all parameter maps. We observed no effect on the registration result using this value.

In addition to the three available parameter maps, we created a fourth map called Par00GM, in an attempt to improve upon the existing maps. The same modifications described above were incorporated into the Par00GM parameter map to ensure comparable registration results. For the Par00GM, the rigid and affine transformation steps were chosen to be similar to the Par0033 configurations. However, the number of resolutions chosen for the rigid transformation was decreased to 2 (from 4 in Par0033) and increased to 4 (from 2 in Par0033) in the affine transformation step.

For the b-spline configuration, we chose a similar configuration to Par0033. In order to accurately delineate the subcortical structures in the STP data, a fine grid spacing control point was set with a spacing of 10 voxels in all three dimensions. For comparison, par0025 had a 25 voxel spacing, par0033 had a 35 voxel spacing, and par0038 had a 40 voxel spacing. A default grid spacing of (8 4 2 1) across four resolutions was chosen, which spaces the grid by an extra factor of 2 for each resolution to encourage large structure registration at the downsampled resolutions and fine structure resolution at the last (original) resolution stage. A grid pyramid schedule was also implemented to force a Gaussian smoothing of  $\sigma=1/2$  for each downsampling step. We chose this pyramid schedule to prevent excessive smoothing and loss of signal information when downsampling the data. A maximum step length of 0.2 mm was implemented to give some additional flexibility to the registration to account for biological variability between samples, without encouraging gross and unrealistic deformations.

A list of all parameters for each parameter configuration can be seen in Supplementary Tables 1-3.

### **Sample preparation for STP imaging**

Test registration mouse brains for STP imaging were extracted following a transcardial perfusion process using 20ml 1X phosphate-buffered saline (PBS) at 4°C, followed by 20ml 4%

paraformaldehyde (PFA) at 4°C. Brain tissues were post-fixed in 4% PFA and kept at 4°C for 24 hours before transferring to 1X PBS with 0.01% NaN<sub>3</sub>. To embed the sample in preparation of imaging, a 4.5% oxidised-agarose solution was prepared by stirring agarose (type 1; Sigma) with 10 mM NaIO<sub>4</sub> (Sigma) and 50 mM phosphate buffer (PB) at room temperature in a light-shielded environment for 2 hr 30 mins. The oxidised-agarose solution was filtered three times with PB before re-suspending in PB and stored at 4°C until performing the embedding process. During embedding, two 25G needles were horizontally inserted 5 mm apart, through a disposable cube-shaped embedding mould (VWR #E6032-1CS) to provide a horizontal platform to support the mouse brain. The oxidised-agarose solution was heated to boiling point in a microwave, then left to cool until 65°C. After filling the mould, the mouse brain was placed onto the platform using a set of forceps. By using the needle platform to support the brain, the sample was aligned within the mould and did not require slicing a flat section from the tissue. This allowed imaging of the entire intact sample. Finally, the embedded sample was left to passively cool to room temperature until it became solid, before carefully extracting from the moulds and immersed and stored within 1X PBS with 0.01% NaN<sub>3</sub> at 4°C until we performed imaging.

### **STP imaging**

Agarose embedded samples were glued with a cyanoacrylate-adhesive (Loctite Ltd.) onto a glass slide, with the anterior-posterior axis of the mouse brain positioned orthogonal to the slide. The base of the slide consisted of two magnetic blocks which allowed the sample to be magnetically fastened to a magnetic base in a custom-designed 1.5L plexiglass water-bath. To maintain tissue integrity over several days of imaging, a custom-designed cooling system (JULABO UK Ltd.) was inserted into the bath and set to maintain a 4°C environment. The bath was then filled with PB and translated to align the sample with the imaging objective and microtome. During image acquisition, physical sections were acquired every 50 µm using the microtome, with five optical sections using a piezo controlled objective acquired per physical section, providing a 10 µm axial resolution. These were combined with a 16x, 0.8NA imaging objective (Nikon Inc.) to obtain a spatial resolution of 0.54 µm. The first optical layer was acquired 25 µm below the surface to avoid surface distortions appearing in the image data. Complete coronal sections of the mouse brains were acquired by tiling the objective 1x1 mm field-of-view across the coronal plane. This resulted in a 12 x 10 tile mosaic for each optical

section. Physical sections were obtained using a microtome with a blade vibration frequency of 60 Hz and velocity 0.35 mm/s. Laser (Chameleon Ultra II, Coherent) excitation was set at 920 nm for GFP excitation. The laser power was incrementally increased for optical section acquisition at deeper depths to compensate for laser power loss induced from scattering. Tissue autofluorescence was collected over three photomultiplier tubes (PMTs) channels (red, green and blue). PMT gain was set to 850V for all three channels. Tile acquisition was conducted with a resolution of 2080 x 2080 pixels with 16-bit range, amounting to around 3 MB per tile. Imaging conducted with these parameters resulted in 250 optical sections along the Z-axis (rostral-caudal axis) with each section comprising 120 tiles. The acquisition occurred over 5 days and generated just over 3 TB of data per brain. All images were saved to a network-attached storage (NAS) drive (DS1815+; Synology Inc.). Following imaging, free-floating slices were extracted from the PB and stored in 1X PBS with 0.01% NaN<sub>3</sub> at 4°C.

### **Data preparation for Elastix**

The Allen CCF3 atlases (average and annotation) are both isotropic 10 µm voxel size. We, therefore, downsampled the STP data to 10 µm isotropic size by scaling the dataset in XY by factor 0.054. The Z-dimension did not need resampling as STP imaging was performed with 10 µm Z-resolution. Using ImageJ/Fiji, the data was converted to an 8-bit range with the pixel intensity scaled to add contrast for autofluorescent structures.

A 3D median filter of kernel size  $k=5$  was applied to the STP data sets and saved alongside the unfiltered versions. This was done to evaluate whether additional image processing steps could assist registration accuracy.

### **Registration with Elastix**

Registration was performed with Elastix using the STP data, average CCF3 atlas and the annotation atlas as the fixed image and two moving images respectively. A binary hemisphere atlas was created by creating a zero-value atlas of the same dimensions as the annotation atlas. Everything left of the midline in the XY plane was denoted left hemisphere, and everything right of the midline was labelled as the right hemisphere by setting the pixel values to 1. This allowed structure distinction between hemispheres. Three parameter maps were initiated in order of rigid, affine and non-rigid using the parameters specified by the

parameter maps (Supplementary Tables 1-3). Registration was performed from average CCF3 to the STP data to provide a spatial mapping from the STP data to the CCF3 atlas.

The transformation result was applied to the annotation CCF3 atlas, the hemisphere atlas, and a set of points in the fixed image domain. When applied to the annotation atlas, nearest neighbour interpolation was used as a resample interpolation method to ensure the labels were not corrupted during the transformation.

All registration with Elastix was executed on an Ubuntu Intel(R) Core(TM) i9-7980XE CPU at 2.60GHz with 32 cores and 128 GB RAM workstation.

### **Quantitative assessment of target registration error**

Elastix registration accuracy was determined by calculating a Euclidean distance TRE between the transformed fixed points set and the original moving points set. Two samples, instead of one, were chosen to give an idea of the stability of parameter performance between data sets.

Two PVCreSox14 brain samples were imaged under the same scan conditions and downsized to 10  $\mu\text{m}$  isotropic voxel size. Sample 1 contained a significant amount of tissue tearing, particularly in the inferior colliculus, ventral hypothalamus and piriform region of the isocortex. These originated from the transcatheter perfusion for brain fixing and extraction. Sample 2 was free from tissue tearing.

During landmark demarcation, the average CCF3 atlas was manually referenced against the two sample atlases to get the shared rostral-caudal extent of all data sets. This step ensured landmarked positions were anatomically present across all data sets. At the most caudal position, coronal slice position 266 at AP -5.4 in the average CCF3 atlas was aligned with coronal slice 1 in Sample 1, and coronal slice 104 in Sample 2. At the most rostral position, coronal slice 810 at AP +0.1 in the average CCF3 atlas was aligned with coronal slice 525 in Sample 1, and coronal slice 660 in Sample 2. Landmarked points were therefore limited to a coronal volume in the range AP -5.4 to AP +0.1. We located a total of 84 corresponding points in the average atlas and both test sample atlases, using ImageJ/Fiji. Positions were chosen

based on global and local anatomical features. Global regions included, for example, the surface of the isocortex and the exposed ventral region of the hypothalamus. Local regions included, for example, marked positions in the LGd, columns of fornix, mammillothalamic tract and lateral ventricles.

Following Elastix registration between both samples and the CCF3 atlas, the resulting transformation map was applied to the fixed points of the sample data to map them to the CCF3 moving image domain. A TRE was calculated by finding the Euclidean distance between all 84 corresponding points (Eq. 3.1).

### **Quantitative assessment of human landmark identification accuracy**

To put the magnitude of the automated Elastix registration error into perspective, we measured a human landmark identification error. A total of 20 slices were extracted from the average CCF3 atlas starting from initial coronal slice position of 6 and incrementing 66 slices until the end. In each slice, 4 positions were manually landmarked at randomly identified anatomical positions, using ImageJ/Fiji, giving a total of 80 positions. Landmarks were chosen based on global and local anatomical features and included positions on the outer surfaces of the brain and sub-cortical positions including white matter tracts and ventricles.

A complete 1320 slice average atlas was presented to a blinded human landmark identifier with access to the reference of 20 landmarked slices. We tasked the human identifier with marking the same positions in the complete unmarked atlas. To not bias the human landmark identifier to group together the marking of 4 positions on the same slice, they were informed that markings grouped together on a single slice in the landmarked reference atlas did not necessarily originate from that same slice. The human identifier, therefore, matched corresponding positions without restriction in the rostral-caudal extent. Human identified landmarked positions were used to give a target error by calculating a Euclidean distance between each corresponding point (Eq. 3.1).

### **Qualitative assessment of Elastix registration accuracy**

Final transformation maps generated through Elastix were applied to the CCF3 annotation atlas for both test sample sets. The segmentation atlases were converted to outlines for each

anatomical region and overlaid over the sample atlases using ImageJ/Fiji. Delineation accuracy was visually assessed by appraising alignment both globally (for example, the cortical surface boundary) and locally (for example, sub-cortical structures including LGd, dentate gyrus, lateral ventricles, and myelinated fibre tracts).

### **Quantitative similarity scores of Elastix registration**

To assess the similarity of the moving average CCF3 registration to the sample 1 and 2 STP data sets, a set of 2048 randomly sampled windows of  $100 \times 100$  voxels in XY were extracted from randomly positioned coronal sections in the STP sample data and the same positions in the resampled average CCF3 data. We calculated the correlation and mutual information for each randomly sampled region in the sample atlases against the same region in the average atlas which should now be registered. We chose to only assess a set of randomly sampled windows instead of using the entire data set for computational efficiency. We imposed a restriction on the valid sampling regions with the constraint that over 75% of the pixels needed to be non-zero. This ensured random samples contained information relating to the tissue rather than to the background. When evaluating the impact of median filtered data on registration result, random sampling was done on the un-filtered data, so the filtering process did not impact the similarity measures.



### 3.3 RESULTS

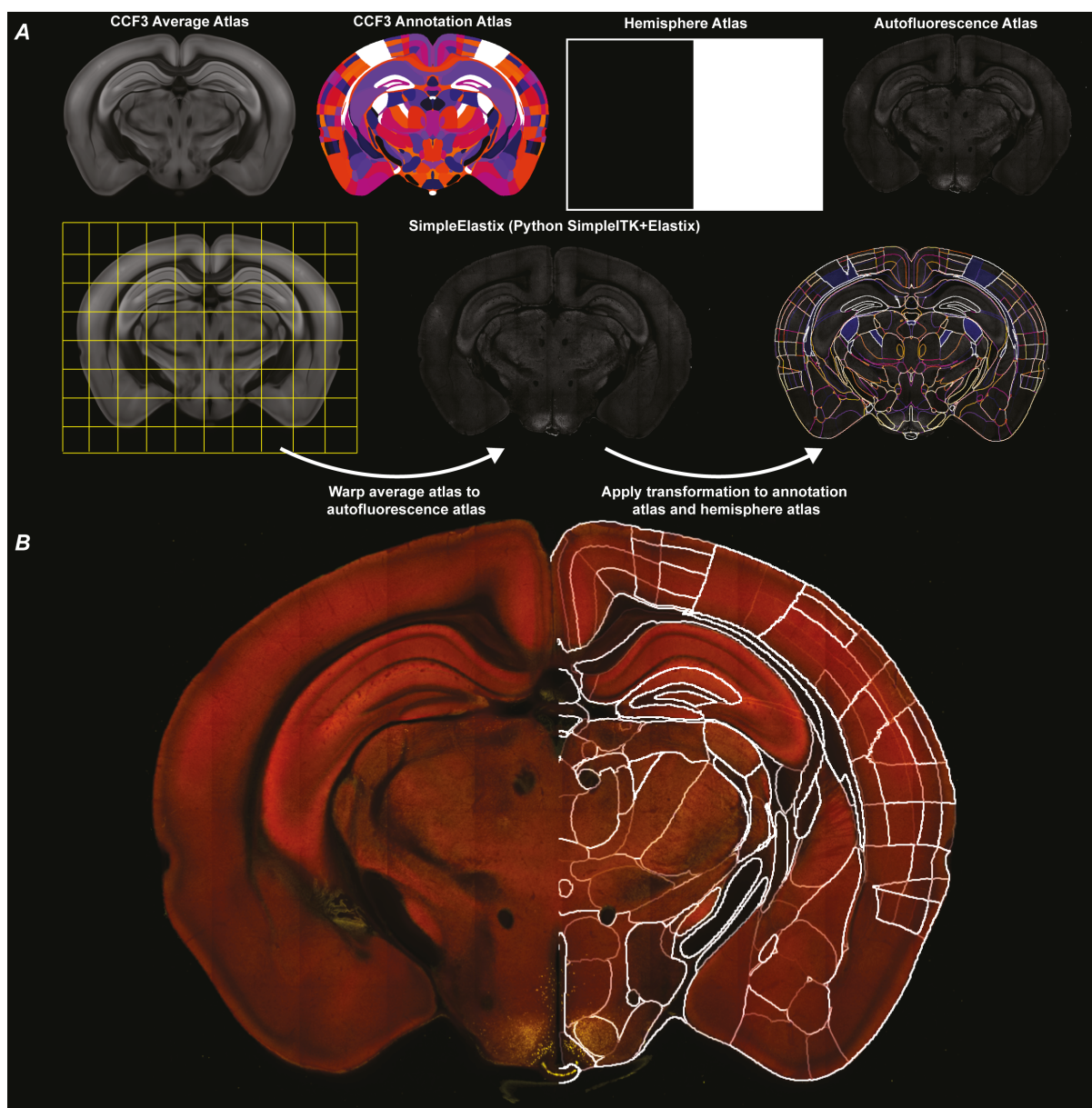
#### **Custom Elastix pipeline enabled STP segmentation and hemisphere distinction**

Elastix was successfully used to register whole mouse brain STP data to the Allen CCF3 common coordinate framework atlas using a set of rigid, affine and non-rigid transformations. The pipeline performed registration between the average CCF3 atlas and the STP sample data sets, with the resulting transformation map applied to the annotation CCF3 atlas. After executing the pipeline, the transformed annotation atlas permitted segmentation and identification of thousands of anatomical regions within the STP data sets (Fig. 3.3A). The combination of rigid, affine and a non-rigid b-spline registration process allowed robust results even considering the whole brain STP data was not originally spatially aligned to the CCF3 atlases. Applying the resulting transformation on the hemisphere atlas allowed the hemispherical distinction of all structures, which was not possible with the CCF3 atlas alone (Fig. 3.3B). Registration took approximately 30 minutes to execute all stages; these were the registration to CCF3 atlas, resampling the annotation atlas, and resampling the hemisphere atlas. The registration to the CCF3 atlas took the majority of the execution time at 20-25 minutes with the resampling of the annotation and hemisphere atlas taking a shorter 5 minutes.

Structures of interest could be located by their identifier number from the Allen Brain Institute hierarchical taxonomy, for which the identifier value matches the pixel intensity in the 32-bit annotation atlas. In addition to locating brain regions, we could also use the identifiers to extract structure masks or perimeters for upscaling and use on the original STP data set resolution.

#### **Quantitative assessment of target registration error (TRE)**

All parameter configurations (Par00GM, Par00GM, Par0033 and Par0038) were tested on the two test samples; brains extracted from a PVCreSox14 mouse line. Sample 1 exhibited significant brain tearing and tissue corruption (Fig. 3.6A). Sample 2 was clear of tearing (Fig. 3.6B). Each parameter map was executed under the same conditions. To assess if median filtering improved registration result, each sample data set was median filtered (kernel size  $k=5$ ). The final transformation map was used to calculate a TRE (Eq. 3.1) of 84 sparsely located points between each sample data set and the average CCF3 atlas (Fig. 3.4).



**Figure 3.3.** Registration pipeline of STP data to Allen CCF3 and segmentation using the annotation atlas. **A.** Registration requires the use of the average atlas, annotation atlas, a generated hemisphere atlas, and the STP autofluorescence data (top row). Registration warps the average atlas using a set of rigid, affine and lastly non-rigid using grid control points, until spatial aligned to the autofluorescence data. The final transformation is applied to the annotation atlas for delineation of thousands of anatomical structures in the STP data. Registration with an STP data set of isotropic voxel size of 10  $\mu\text{m}$  takes between 30 and 45 minutes. **B.** The annotation atlas (white outlines) can be projected over the STP data (red) for visual assessment. The hemisphere atlas also enables distinction of structures between the hemispheres.

When no median filtering process was used on the STP data sets prior to registration, the TRE for sample 1 was calculated at  $17.29 \pm 1.33$  voxels (Mean  $\pm$  SEM, n=84 points) for Par00GM,  $18.83 \pm 1.37$  voxels for Par0025,  $17.63 \pm 1.27$  voxels for Par0033 and  $19.58 \pm 1.48$  voxels for Par0038 (Fig. 3.4B). The mean TRE for sample 2 was  $19.54 \pm 1.51$  voxels for Par00GM,  $20.29 \pm 1.50$  voxels for Par0025,  $19.77 \pm 1.48$  voxels for Par0033 and  $22.25 \pm 1.66$  voxels for Par0038. In equivalent dimensional units using the isotropic voxel size of  $10 \mu\text{m}$ , the best accuracy for sample 1 was  $172.9 \pm 13.3 \mu\text{m}$  using the Par00GM configuration and  $195.4 \pm 15.1 \mu\text{m}$  for sample 2 also using the Par00GM configuration. Par0038 consistently performed worst for both samples with Par0033 and Par0025 coming second and third, respectively.

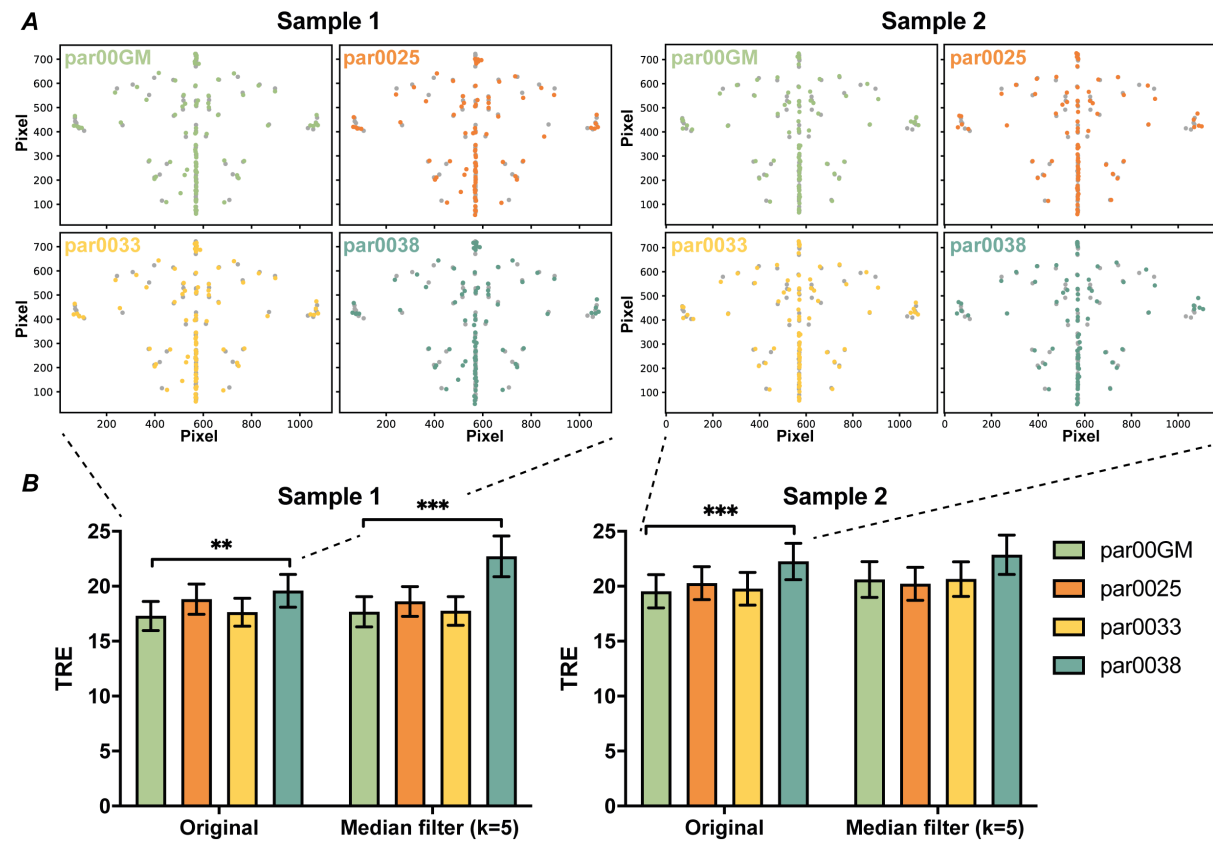
Significance tests of one-way ANOVA using matched rows, assuming non-equal variability of differences with a Geisser-Greenhouse correction, showed significant differences between parameter configuration TRE results with  $p=0.0425$  (n=84 points) for sample 1 and  $p=0.0001$  (n=84 points) for sample 2. Multiple comparisons between each parameter map showed significant improvement with the Par00GM configuration over Par0038 for sample 1 ( $p=0.0012$ , n=84 points), and significant improvement of Par00GM over Par0038 for sample 2 ( $p=0.0005$ , n=84 points).

For the median filtered data, the TRE was almost universally lower by  $74 \pm 38 \mu\text{m}$  (Mean  $\pm$  SEM, n=4 parameter configurations) across all parameter configurations. However, Par00GM saw an increase in TRE of  $2.1 \mu\text{m}$  and  $0.6 \mu\text{m}$  for sample 1 and 2, respectively, when compared to their unfiltered test results.

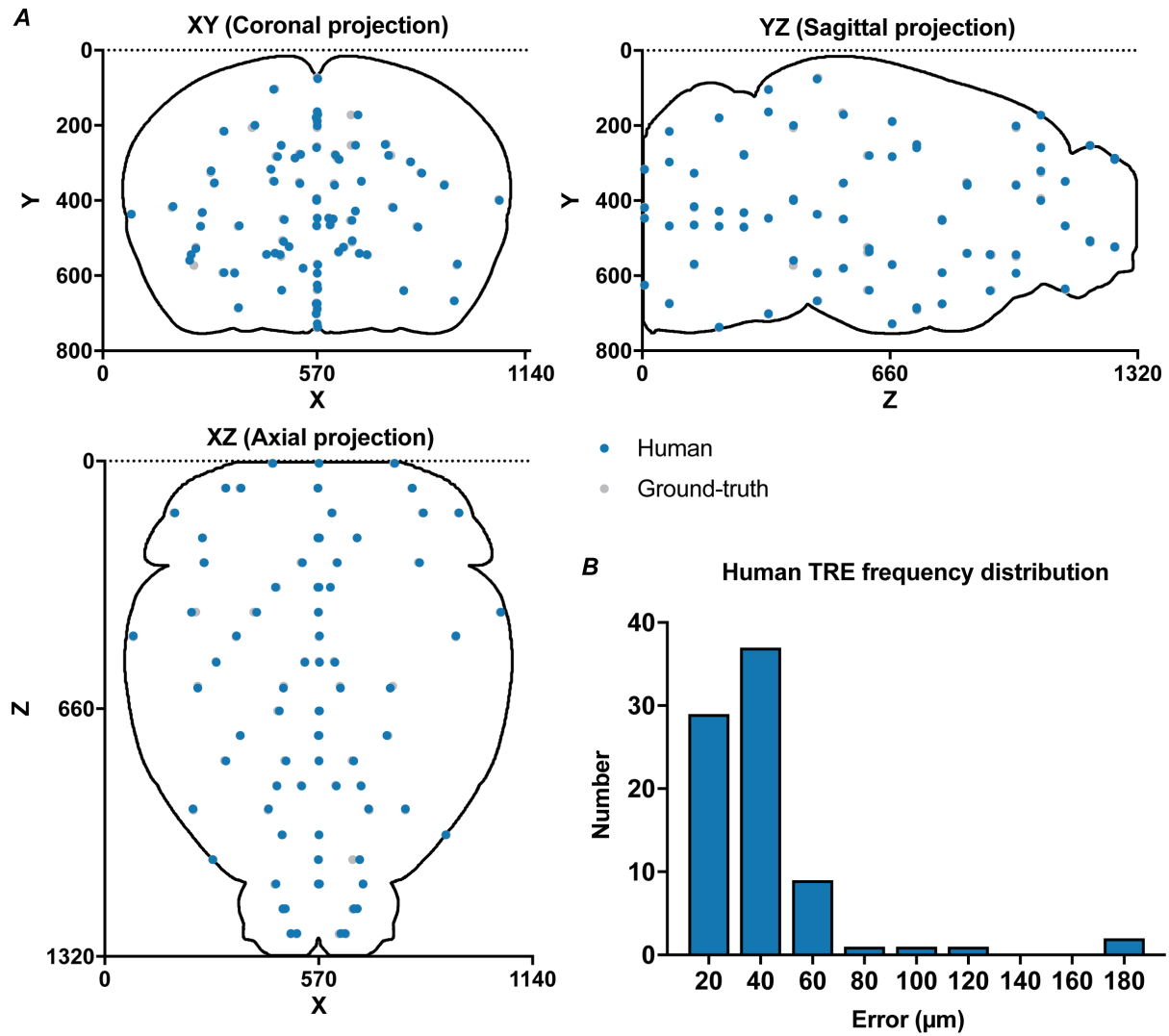
Significance tests of one-way ANOVA using matched rows on the median filtered data, assuming non-equal variability of differences with a Geisser-Greenhouse correction, showed significant differences in parameter configuration TRE for sample 1 with  $p=0.0001$  (n=84 points) and for sample 2 with  $p=0.0042$  (n=84 points). Multiple comparisons between each parameter map showed significant improvement with the Par00GM configuration over Par0038 ( $p=0.0002$ , n=84 points) for sample 1.

The Par00GM configuration performed the best overall when quantitatively assessed for the four test cases (samples 1 and 2 with original and median filtered cases) for TRE.

All parameter maps executed in equivalent times of approximately 30 minutes.



**Figure 3.4.** Transformation of sparse points for target registration error (TRE) evaluation in the four tested parameter map configurations. **A.** Sparsely labelled points ( $n=84$ ) in the fixed STP images were transformed into the moving CCF3 image domain using the final transformation calculated during the registration process. Plots are for the unfiltered STP cases and show the moving points (dark grey) and the transformed fixed points (colour coded to the key in panel B). **B.** Evaluation of TRE scores for the original unfiltered tests and the median filtered ( $k=5$ ) tests for samples 1 and 2. Bars represent mean  $\pm$  SEM ( $n=84$  points). Par00GM configuration performed best with significance  $p<0.01$  (\*\*) and  $p<0.001$  (\*\*\*) over Par0038 for sample 1 and  $p<0.001$  (\*\*\*) over Par0038 for sample 2. Significance test of multiple comparisons with one-way ANOVA using matched rows, assuming non-equal variability of differences with a Geisser-Greenhouse correction.



**Figure 3.5.** Human target error assessment. **A.** Coronal, sagittal and axial projections of ground-truth landmark positions ( $n=80$ ) (grey markers) and human identified corresponding position (blue markers). Respective projection atlas outlines added for reference. **B.** Histogram of human target error. Human target error of  $38.3 \pm 3.3 \mu\text{m}$  (Mean  $\pm$  SEM,  $n=80$  points).

### Quantitative Assessment of Human Landmark Identification Accuracy

To measure a human error in landmark identification, 80 landmarked points in the average CCF3 atlas were identified by a human performer (Fig. 3.5). The human performer identified corresponding points with an error of  $38.3 \pm 3.3 \mu\text{m}$  (Mean  $\pm$  SEM,  $n=80$  points). Errors were mostly confined to the XY plane (Fig. 3.5A), with the correct Z-plane identified for most cases. The error distribution between ground-truth and human landmarks were constrained to low values with a median TRE of  $40 \mu\text{m}$  (Fig. 3.5B). Only a handful of demarcations were made with an error over  $80 \mu\text{m}$ . These errors were mostly limited to the same XY plane, suggesting

the positional regions in XY are harder to identify than in Z. Considering the reference landmarks were presented to the human identifier in the coronal view which contains more contextual anatomical information in XY, this result is not too surprising. However, this observation does present a limitation of error measures during comparison with the Elastix TRE.

### **Qualitative assessment of parameter maps**

Final transformations maps generated by Elastix were applied to the annotation CCF3 atlas. The resampled annotation atlases were converted into outlines and projected onto the STP data sets (Fig. 3.6). The correspondence of delineation outline with autofluorescence was visually assessed across the complete STP data set. The STP data sets contain several hundred individual coronal sections, so for visual representation purposes we show only two slices for comparison for samples 1 and 2 and for the non-median filtered data, only.

For sample 1, slice sections AP -2.7 and AP +0.7 were chosen for representative comparison across the parameter map configurations. Slice AP -2.7 is a coronal cross-section centrally positioned in the anterior-posterior axis, showing sensory-motor and polymodal thalamic areas, Ammon's horn and dentate gyrus regions of the hippocampus, all zones of the hypothalamus and also isocortex. Slice AP +0.7 is a coronal cross-section in the anterior region of the brain showing striatum and pallidum regions of the cerebral nuclei, and isocortex areas. For sample 2, representative slices were chosen at AP -4.4 and AP -2.6 for comparison across all parameter maps. Slice AP -4.4 is positioned posterior in the brain and shows isocortex and retrohippocampal regions, both sensory and motor midbrain regions and also areas of the pons in the hindbrain. Slice AP -2.6 is a coronal section positioned centrally in the anterior-posterior axis, containing regions equivalent to AP -2.7.

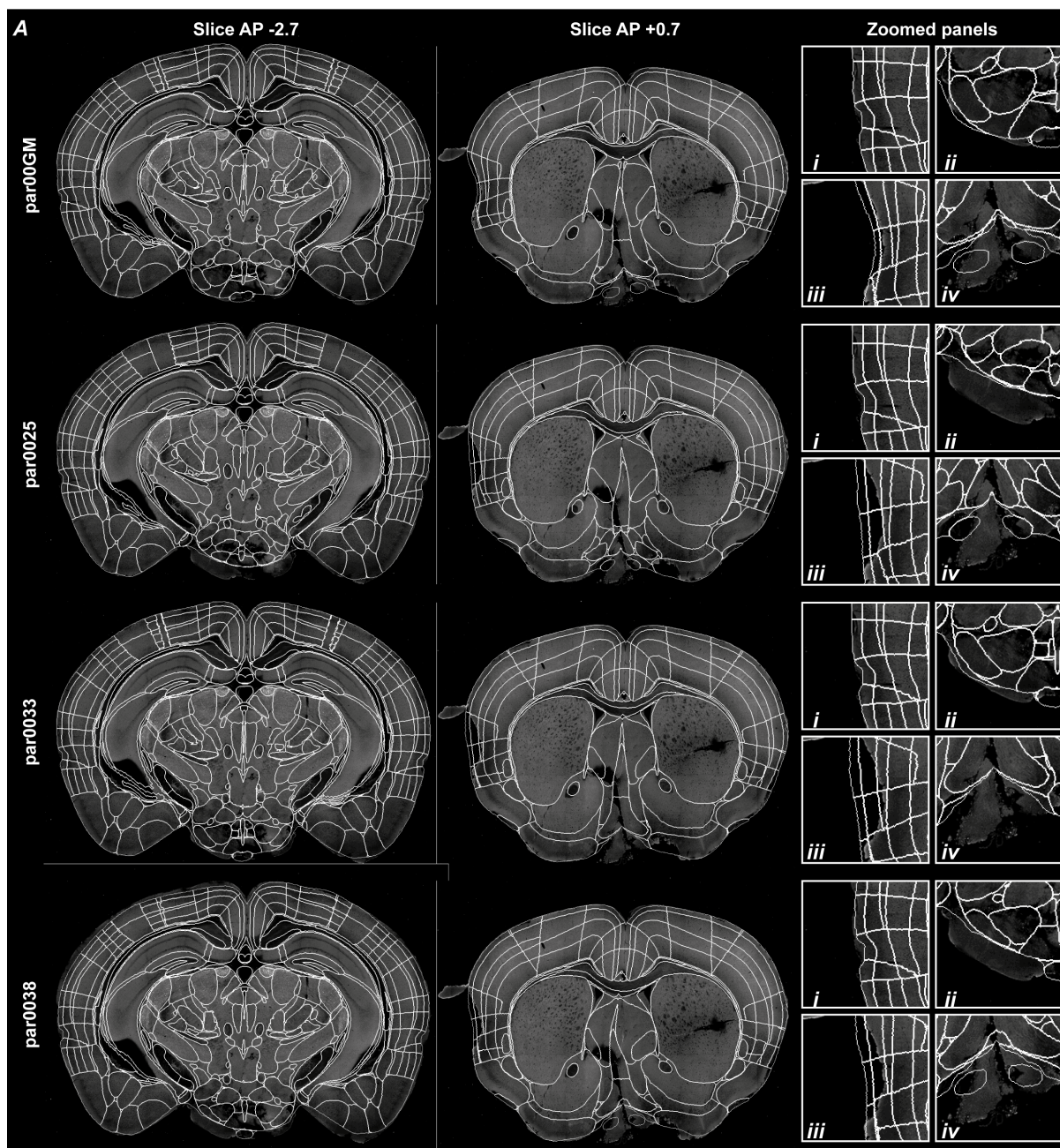
Visual appraisal of the registration accuracy demonstrates the limitation of TRE based analysis which fails to reflect the accuracy of registration comprehensively. For sample 1 (Fig. 3.6A), the ventral surface of the hypothalamus in slice AP -2.7 did not align well in all but the Par00GM parameter configuration (Fig. 3.6Aii). Par0033 was closer in alignment whilst Par0025 and Par0038 performed poorly. Par0038 also failed to align the olfactory area in the most dorsal-lateral region of the isocortex in the right hemisphere of slice AP -2.7.

In slice AP +0.7, all parameter maps performed similarly in global alignment. However, the torn section of tissue in the isocortex of the left hemisphere was handled differently by each parameter configuration. Par00GM and Par0038 attempted to align with the exposed tissue, whilst Par0025 and Par0033 do not (Fig. 3.6Aiii). Furthermore, the midline appeared to deform in all but the Par0033 configuration when attempting to compensate for the shearing of the tissue in the dorsal-medial axis. The ventral periventricular hypothalamus region of slice AP +0.7 (Fig. 3.6Aiv) was grossly mismatched, with Par00GM performing the best. The Par00GM configuration also warped the midline in correspondence with the anatomical midline. Overall, Par00GM appeared to perform the best, with Par0033 coming a close second. Both Par0025 and Par0038 performed poorly.

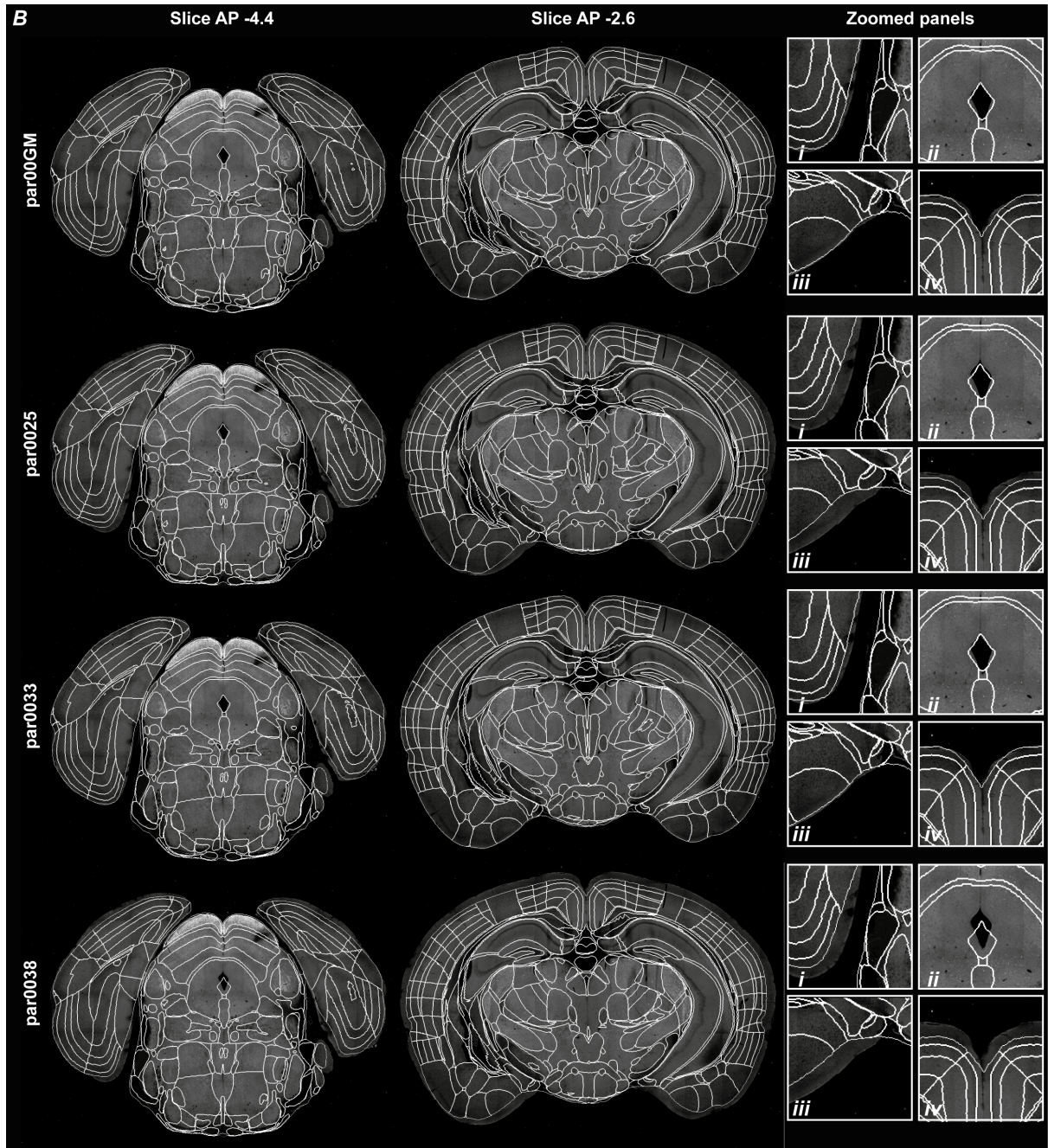
For sample 2 position AP -4.4 only Par00GM and Par0033 accurately mapped the outline of the isocortex, with Par00GM performing optimally (Fig. 3.6Bi). The cerebral aqueduct (Fig. 3.6Bii) was registered well in all but Par0038 which misaligned the position by over 10 voxels, or equivalently 100  $\mu\text{m}$ . More accurate alignment was seen in the Par00GM and Par0033 configurations.

In slice AP -2.6, the most significant indication of parameter alignment success could be seen in the ventrally positioned cortical amygdala region (Fig. 3.6Biii). There was an apparent misalignment in the Par0025 and Par0038 configuration with Par00GM and Par0033 performing well. The most dorsal positioned cortical region was also misaligned in Par0025 and Par0038. Par00GM and Par0033 performed equally well.









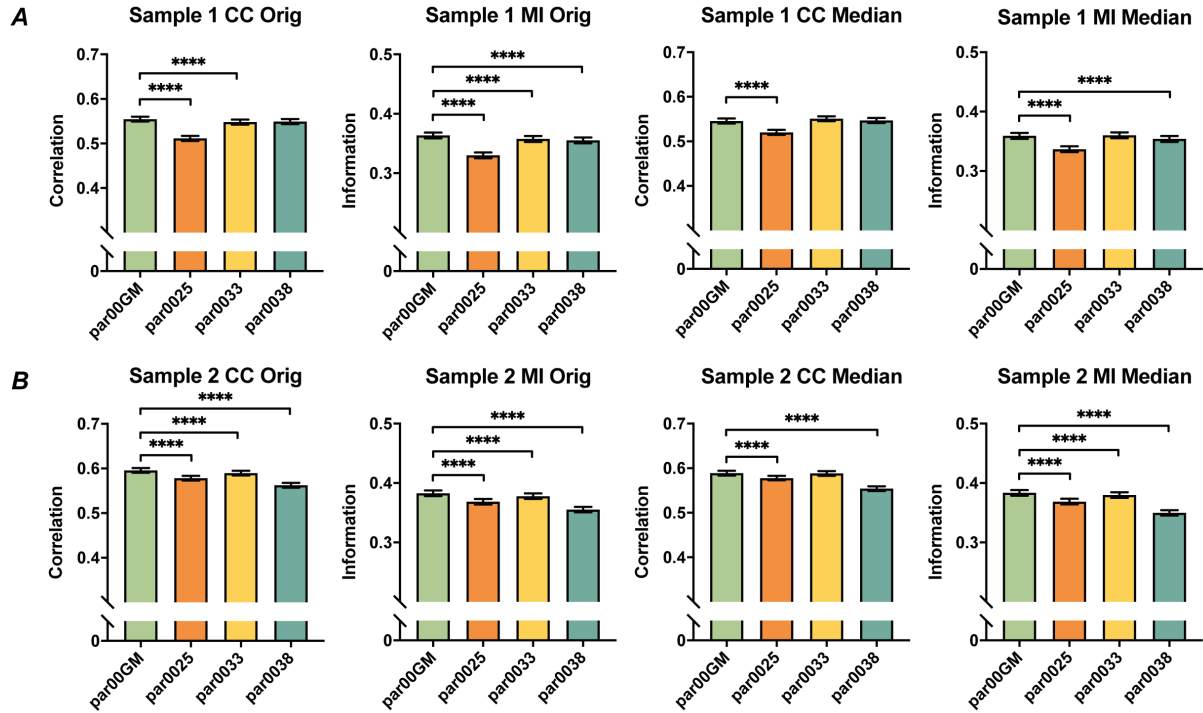
**Figure 3.6.** Qualitative registration result. **A.** Two respective slice sections AP -2.7 and AP +0.7 in sample 1 unfiltered cases. Tissue corruption can be seen in slice AP +0.7. Par00GM and par0033 perform comparatively well, at global alignment (ii) and particularly when aligning to the corrupted tissue areas (iii, iv). **B.** Two respective slice sections AP -4.4 and AP -2.6 in sample 2 unfiltered cases. Par00GM and par0033 again perform well with good alignment of local structures such as the cerebral aqueduct (ii) and ventral alignment around the cortical amygdala region (iii).

### **Quantitative similarity scores of Elastix registration**

The correlation coefficient (CC) and mutual information (MI) similarity measures between randomly sampled regions ( $n=2048$ ) of size  $100 \times 100 \times 1$  voxels ( $1 \text{ mm} \times 1 \text{ mm} \times 10 \text{ }\mu\text{m}$  in dimensional units) in sample 1 and 2 showed highest image similarity measures for the Par00GM parameter configuration (Fig. 3.7).

Multiple comparisons with one-way ANOVA using matched rows, assuming non-equal variability of differences with a Geisser-Greenhouse correction, showed significant differences between all parameter configurations for the correlation coefficient and mutual information measures for the original and median-filtered data;  $p<0.0001$  ( $n=2048$  randomly sampled regions) for sample 1 correlation coefficient on unfiltered data and  $p<0.0001$  for the median filtered data;  $p<0.0001$  for sample 2 correlation coefficient on unfiltered data and  $p<0.0001$  for the median filtered data;  $p<0.0001$  for sample 1 mutual information measure on the unfiltered data and  $p<0.0001$  for the median filtered data;  $p<0.0001$  for sample 2 mutual information measure on unfiltered data and  $p<0.0001$  for the median filtered data. Multiple comparisons between parameter configurations for all test cases (unfiltered data, median filtered data, correlation coefficient measure, and mutual information measure) showed the Par00GM configuration to have significant improvement of  $p<0.0001$  ( $n=2048$  randomly sampled regions) in nearly all comparisons with the other configurations.

We, therefore, observed optimum performance with the Par00GM configuration. The Par0025 and Par0038 configurations were consistently the worst performers in all scenarios with Par0038 performing better than Par0025 when used to sample 1 which had tissue tearing, and Par0025 performed better than Par0038 when used to register sample 2 with no tearing.



**Figure 3.7.** Similarity measures by correlation coefficient (CC) and mutual information (MI). Measures were taken for randomly sampled regions ( $n=2048$ ) of size  $100 \times 100 \times 1$  voxels ( $1 \text{ mm} \times 1 \text{ mm} \times 10 \mu\text{m}$  in dimensional units) between STP and transformed average CCF3 atlas. **A.** Sample 1 original and median filtered STP cases show significantly higher similarity measures for Par00GM  $p < 0.0001$  (\*\*\*\*) for all indicated comparisons. Bars represent mean  $\pm$  SEM ( $n=2048$  randomly sampled regions). **B.** Sample 2 original and median filtered STP cases also show significantly higher similarity measures for Par00GM  $p < 0.0001$  (\*\*\*\*) for all indicated comparisons. Bars represent mean  $\pm$  SEM ( $n=2048$  randomly sampled regions). Significance testing was done using multiple comparisons with one-way ANOVA using matched rows, assuming non-equal variability of differences with a Geisser-Greenhouse correction.

### 3.4 DISCUSSION

Of the four parameter maps tested, Par00GM and Par0033 performed the best in both sample 1 and sample 2 cases for all three quantitative and qualitative test cases. In qualitative testing, both parameter configurations handled the torn tissue sections (Fig. 3.6Aiii) by either attempting to align, as seen with Par00GM, or by ignoring the deformation and maintaining the same global outline in the absence of any data, as seen in Par0033. However, Par0033 did not warp the midline as prominently, when compensating for the tissue shearing in the dorsal-medial direction (Fig. 3.6iv). From all testing scenarios, Par0025 and Par0038 did not perform well, exhibiting gross global misalignment particularly in ventral regions of the hypothalamus and cortical regions. This was a surprising result considering these are the regions with high contrast between tissue and empty background signal in the STP data.

When assessing quantitative TRE accuracy, both Par00GM and Par0033 excelled, with Par00GM performing best with a TRE of  $172.9 \pm 13.3 \mu\text{m}$  (mean  $\pm$  SEM). However, the utility of this test could be limited on account of a sparse number of corresponding points between data sets. In addition, a human target error of  $38.3 \pm 3.3 \mu\text{m}$  (mean  $\pm$  SEM) does suggest human visual assessment is more accurate, although limited as a direct comparison on account that the human performer samples more XY spatial information during the task than the Elastix registration algorithm does during the similarity measures. However, considering the approximate diameter of a neuron soma is  $25 \mu\text{m}$ , a  $170 \mu\text{m}$  Elastix TRE is equivalent to six or seven neuron soma which is remarkably accurate considering the process is completely automated, and with no human guidance. The human annotation error illustrated by Niedworok et al. (2016) (Fig. 3.1) shows that human annotation could be disagreeable from the true annotation by as high as 1 mm in the case for the VPM. If performing manual registration across all three-dimensions, the error in delineation could be significantly higher.

Interestingly, median filtering (kernel size  $k=5$ ) on the STP data prior to registration did not improve the registration accuracy. This result was surprising given that the median filter process is edge-preserving and therefore should increase the contrast between anatomical structures. It is possible that the process of median filtering introduced more zero value pixels within the tissue. A bilateral filtering process could provide more robust filtering, which is edge-preserving. A difference of Gaussians is often used for feature enhancement in images

and could modify the intensity profile of the fixed and moving images which would assist the registration process.

Given the high similarity scores, TRE and visual alignment of the Par00GM parameter configuration, it was decided Par00GM should be the registration parameter configuration used for future whole-brain registration applications. The finer grid control point scheme aided registration even when the sample was suffering from tissue tearing. This means the Par00GM parameter configuration would be more robust for handling a variety of global and local anatomical structures and any biological variability between the fixed and moving data. Whilst Par0033 performed admirably, it did not perform optimally for finer alignment, especially in the presence of inhomogeneous tissue structure (Fig. 3.6A), for which Par00GM did perform well. Ultimately, the use of the registration process was to provide accurate delineation for all structures in the brain (at the limitation of the atlas being used), so downstream analysis can provide reliable results with minimal error originating from the registration technique. Par00GM provides the best configuration for this, allowing thousands of anatomical brain regions to be identified in whole mouse brain STP data sets.

### 3.5 FUTURE PERSPECTIVES

TRE analysis is not comprehensive of registration result, with qualitative visual assessment providing a more thorough inspection of the accuracy. In part, this is due to the challenge of identifying a large number of landmarks between multiple data sets. Therefore, an exciting future direction is the exploration of automated identification and matching of landmark features between data sets. This would permit a much larger number of identified landmarks between the fixed and moving data, and a more comprehensive TRE analysis, therein. Additionally, a database of pre-landmarked data sets would provide a more comparable demonstration of registration accuracy for studies which improve on parameter configuration, or studies which develop a novel registration approach.

Algorithms for automated landmark detection already exist, such as the commonly utilised scale-invariant feature transform (SIFT) algorithm, which has been frequently used for feature matching in image stitching applications (Lowe, 2004; Min et al., 2012; Qu et al., 2015; Xing & Miao, 2007). As a scale-invariant approach, features between fixed and moving images which are not of the same scale are detectable as corresponding key points. The challenge for this approach for STP data is that a single 2D image in the dataset does not have a direct correspondence with a single 2D image in the atlas due to misalignment between samples. Any exploration in this area would, therefore, need to account for the three-dimensionality of the STP data.

In the pipeline presented here, we performed registration using 10  $\mu\text{m}$  isotropic data for the fixed STP data and the CCF3 atlases. Faster registration could be achieved with a further downsampled data set, such as  $25 \times 25 \times 10 \mu\text{m}$ . However, downsampling would remove some spatial features in the data so additional care would be required to ensure end result accuracy. Another study could look at the robustness of registration result for different image filtering processes. Whilst we only evaluated the median filter process here, Gaussian filtering, bilateral filtering or using a difference of two Gaussians could help enhance features and therefore Elastix registration result.

Another direction could be the exploration of different interpolation algorithms to help improve sampling in the Z-axis for the YSK atlas (Chon et al., 2019). As the YSK atlas offers the

most up-to-date delineation and labelling scheme, a more comprehensive rostral-caudal sampling would make it better suited as a common coordinate framework for use with whole-brain data sets. Interpolation could be achieved by using distance maps on each unique masked region, or through a manual assisted to ensure regions are not corrupted during interpolation.

Lastly, the registration algorithm we used in the pipeline was Elastix on account of computational efficiency and speed. Another established registration algorithm is ANTs which has also been used primarily for MRI data. ANTs also uses pixel intensities and image similarity measures to compute the transformation during registration. However, the algorithm could provide more accurate results and is worth exploring as another approach.

### 3.6 SUPPLEMENTARY DATA

| Parameter                                 | par0025rigid.txt                     | par0033rigid.txt                     | par0038rigid.txt                     | par00GMrigid.txt                     |
|-------------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| Automatic Scales Estimation               | TRUE                                 | TRUE                                 | TRUE                                 | TRUE                                 |
| Automatic Transform Initialisation        | TRUE                                 | TRUE                                 | TRUE                                 | TRUE                                 |
| Automatic Transform Initialisation Method | Center Of Gravity                    | Center Of Gravity                    | Center Of Gravity                    | Center Of Gravity                    |
| B Spline Interpolation Order              | 3                                    | 1                                    | 1                                    | 1                                    |
| Default Pixel Value                       | 0                                    | 0                                    | 0                                    | 0                                    |
| Element Spacing                           | 0.01 0.01 0.01                       | 0.01 0.01 0.01                       | 0.01 0.01 0.01                       | 0.01 0.01 0.01                       |
| Erode Mask                                | FALSE                                | FALSE                                | FALSE                                | FALSE                                |
| Final B Spline Interpolation Order        | 3                                    | 3                                    | 3                                    | 3                                    |
| Fixed Image Dimension                     | 3                                    | 3                                    | 3                                    | 3                                    |
| Fixed Image Pyramid                       | Fixed Recursive Image Pyramid        | Fixed Recursive Image Pyramid        | Fixed Recursive Image Pyramid        | Fixed Recursive Image Pyramid        |
| Fixed Image Pyramid Schedule              |                                      |                                      |                                      | 1 1 1 1 1 1                          |
| Fixed Internal Image Pixel Type           | Float                                | Float                                | Float                                | Float                                |
| How To Combine Transforms                 | Compose                              | Compose                              | Compose                              | Compose                              |
| Image Pyramid Schedule                    | 2 2 0                                | 1 1 1 1 1 1 1 1 1 1 1                |                                      |                                      |
| Image Sampler                             | Random                               | Random Coordinate                    | Random                               | Random Coordinate                    |
| Interpolator                              | B Spline Interpolator                | B Spline Interpolator                | B Spline Interpolator                | B Spline Interpolator                |
| Maximum Number Of Iterations              | 2000                                 | 1000                                 | 750                                  | 1000                                 |
| Maximum Step Length                       |                                      | 0.2                                  | 0.145833                             | 0.2                                  |
| Metric                                    | Advanced Normalized Correlation      | Advanced Mattes Mutual Information   | Advanced Normalized Correlation      | Advanced Mattes Mutual Information   |
| Moving Image Dimension                    | 3                                    | 3                                    | 3                                    | 3                                    |
| Moving Image Pyramid                      | Moving Recursive Image Pyramid       | Moving Recursive Image Pyramid       | Moving Recursive Image Pyramid       | Moving Recursive Image Pyramid       |
| Moving Image Pyramid Schedule             |                                      |                                      |                                      | 1 1 1 1 1 1                          |
| Moving Internal Image Pixel Type          | Float                                | Float                                | Float                                | Float                                |
| New Samples Every Iteration               | TRUE                                 | TRUE                                 | TRUE                                 | TRUE                                 |
| Number Of Histogram Bins                  | 32                                   | 32                                   | 32                                   | 32                                   |
| Number Of Resolutions                     | 1                                    | 4                                    | 2                                    | 2                                    |
| Number Of Spatial Samples                 | 2048                                 | 2048                                 | 2048                                 | 2048                                 |
| Optimiser                                 | Adaptive Stochastic Gradient Descent | Adaptive Stochastic Gradient Descent | Adaptive Stochastic Gradient Descent | Adaptive Stochastic Gradient Descent |
| Registration                              | Multi Resolution Registration        | Multi Resolution Registration        | Multi Resolution Registration        | Multi Resolution Registration        |
| Required Ratio Of Valid Samples           | 0.001                                | 0.001                                | 0.001                                | 0.1                                  |
| Resample Interpolator                     | Final B Spline Interpolator          | Final B Spline Interpolator          | Final B Spline Interpolator          | Final B Spline Interpolator          |
| Resampler                                 | Default Resampler                    | Default Resampler                    | Default Resampler                    | Default Resampler                    |
| Result Image Format                       | Tif                                  | Tif                                  | Tif                                  | Tif                                  |
| Result Image Pixel Type                   | Float                                | Float                                | Float                                | Float                                |
| Transform                                 | Euler Transform                      | Euler Transform                      | Euler Transform                      | Euler Transform                      |
| Use Direction Cosines                     | TRUE                                 | TRUE                                 | TRUE                                 | TRUE                                 |
| Write Result Image                        | FALSE                                | FALSE                                | FALSE                                | FALSE                                |



Write Result Image After  
Each Resolution

False False False False

False False False False

**Supplementary table 1.** Rigid transformation parameter configurations for Par00GM, Par0025, Par0033 and Par0038. These parameters are explicitly specified in the parameter map files. Any parameter not specified is set to a default value at Elastix runtime. See Elastix manual for information on default values.

| Parameter                                 | par0025affine.txt                    | par0033affine.txt                    | par0038affine.txt                    | par00GMaffine.txt                    |
|-------------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| Automatic Scales Estimation               | TRUE                                 | TRUE                                 | TRUE                                 | TRUE                                 |
| Automatic Transform Initialisation        | TRUE                                 | TRUE                                 | TRUE                                 | TRUE                                 |
| Automatic Transform Initialisation Method | Center Of Gravity                    | Center Of Gravity                    | Center Of Gravity                    | Center Of Gravity                    |
| B Spline Interpolation Order              | 3                                    | 1                                    | 1                                    | 1                                    |
| Default Pixel Value                       | 0                                    | 0                                    | 0                                    | 0                                    |
| Element Spacing                           | 0.01 0.01 0.01                       | 0.01 0.01 0.01                       | 0.01 0.01 0.01                       | 0.01 0.01 0.01                       |
| Erode Mask                                | FALSE                                | FALSE                                | FALSE                                | FALSE                                |
| Final B Spline Interpolation Order        | 3                                    | 3                                    | 3                                    | 3                                    |
| Fixed Image Dimension                     | 3                                    | 3                                    | 3                                    | 3                                    |
| Fixed Image Pyramid                       | Fixed Recursive Image Pyramid        | Fixed Recursive Image Pyramid        | Fixed Smoothing Image Pyramid        | Fixed Recursive Image Pyramid        |
| Fixed Image Pyramid Schedule              | 1 1 0                                |                                      |                                      |                                      |
| Fixed Internal Image Pixel Type           | Float                                | Float                                | Float                                | Float                                |
| How To Combine Transforms                 | Compose                              | Compose                              | Compose                              | Compose                              |
| Image Pyramid Schedule                    |                                      | 1 1 1 1 1 1 1 1 1 1                  | 2 2 2 1 1 1                          | 1 1 1 1 1 1 1 1 1 1 1 1              |
| Image Sampler                             | Random                               | Random Coordinate                    | Random                               | Random Coordinate                    |
| Interpolator                              | B Spline Interpolator                | B Spline Interpolator                | B Spline Interpolator                | B Spline Interpolator                |
| Maximum Number Of Iterations              | 500                                  | 1000                                 | 2000                                 | 1000                                 |
| Maximum Number Of Sampling Attempts       | 10                                   |                                      |                                      | 10                                   |
| Maximum Step Length                       |                                      | 0.2                                  | 0.145833                             | 0.01                                 |
| Metric                                    | Advanced Normalized Correlation      | Advanced Mattes Mutual Information   | Advanced Mattes Mutual Information   | Advanced Mattes Mutual Information   |
| Moving Image Dimension                    | 3                                    | 3                                    | 3                                    | 3                                    |
| Moving Image Pyramid                      | Moving Recursive Image Pyramid       | Moving Recursive Image Pyramid       | Moving Smoothing Image Pyramid       | Moving Recursive Image Pyramid       |
| Moving Image Pyramid Schedule             | 0 0 0                                |                                      |                                      |                                      |
| Moving Internal Image Pixel Type          | Float                                | Float                                | Float                                | Float                                |
| New Samples Every Iteration               | TRUE                                 | TRUE                                 | TRUE                                 | TRUE                                 |
| Number Of Histogram Bins                  | 32                                   | 32                                   | 32                                   | 32                                   |
| Number Of Resolutions                     | 1                                    | 2                                    | 2                                    | 4                                    |
| Number Of Spatial Samples                 | 2048                                 | 2048                                 | 2048                                 | 2048                                 |
| Optimiser                                 | Adaptive Stochastic Gradient Descent | Adaptive Stochastic Gradient Descent | Adaptive Stochastic Gradient Descent | Adaptive Stochastic Gradient Descent |

|                                                    | Multi<br>Registration          | Resolution | Multi<br>Registration          | Resolution | Multi<br>Registration          | Resolution | Multi<br>Registration          | Resolution |
|----------------------------------------------------|--------------------------------|------------|--------------------------------|------------|--------------------------------|------------|--------------------------------|------------|
| Registration<br>Required Ratio Of Valid<br>Samples | 0.001<br>Final B Spline        |            | 0.001<br>Final B Spline        |            | 0.001<br>Final B Spline        |            | 0.1<br>Final B Spline          |            |
| Resample Interpolator                              | Final B Spline<br>Interpolator |            | Final B Spline<br>Interpolator |            | Final B Spline<br>Interpolator |            | Final B Spline<br>Interpolator |            |
| Resampler                                          | Default Resampler              |            | Default Resampler              |            | Default Resampler              |            | Default Resampler              |            |
| Result Image Format                                | Tif                            |            | Tif                            |            | Tif                            |            | Tif                            |            |
| Result Image Pixel Type                            | Float                          |            | Float                          |            | Float                          |            | Float                          |            |
| Transform                                          | Affine Transform               |            | Similarity Transform           |            | Affine Transform               |            | Affine Transform               |            |
| Use Direction Cosines                              | TRUE                           |            | TRUE                           |            | TRUE                           |            | TRUE                           |            |
| Write Result Image                                 | FALSE                          |            | FALSE                          |            | FALSE                          |            | FALSE                          |            |
| Write Result Image After<br>Each Resolution        | FALSE                          |            |                                |            | False False                    |            | FALSE                          |            |

**Supplementary table 2.** Affine transformation parameter configurations for Par00GM, Par0025, Par0033 and Par0038. These parameters are explicitly specified in the parameter map files. Any parameter not specified is set to a default value at Elastix runtime. See Elastix manual for information on default values.

| Parameter                                 | par0025bspline.txt            | par0033bspline.txt            | par0038bspline.txt            | par00GMbspline.txt            |
|-------------------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| Automatic Scales Estimation               |                               |                               |                               | TRUE                          |
| Automatic Transform Initialisation        | TRUE                          | TRUE                          | TRUE                          | TRUE                          |
| Automatic Transform Initialisation Method | Center Of Gravity             | Center Of Gravity             | Center Of Gravity             | Center Of Gravity             |
| B Spline Interpolation Order              | 1                             | 1                             | 1                             | 3                             |
| Default Pixel Value                       | 0                             | 0                             | 0                             | 0                             |
| Element Spacing                           | 0.01 0.01 0.01                | 0.01 0.01 0.01                | 0.01 0.01 0.01                | 0.01 0.01 0.01                |
| Erode Mask                                | FALSE                         | FALSE                         | FALSE                         | FALSE                         |
| Final B Spline Interpolation Order        | 3                             | 3                             | 3                             | 3                             |
| Final Grid Spacing In Voxels              | 25 25 50                      | 35 35 35                      | 40 40 40                      | 10 10 10                      |
| Fixed Image Dimension                     | 3                             | 3                             | 3                             | 3                             |
| Fixed Image Pyramid                       | Fixed Smoothing Image Pyramid | Fixed Smoothing Image Pyramid | Fixed Recursive Image Pyramid | Fixed Smoothing Image Pyramid |
| Fixed Image Pyramid Schedule              | 0 0 0 0 0 0 0 0 0 0           | 0 0 0 0 0 0 0 0 0 0           |                               | 1 1 1 1 1 1 1 1 1 1           |
| Fixed Internal Image Pixel Type           | Float                         | Float                         | Float                         | Float                         |
| Grid Spacing Schedule                     | 1 1 1 1 1 1                   | 8 4 2 1                       |                               |                               |
| How To Combine Transforms                 | Compose                       | Compose                       | Compose                       | Compose                       |

|                                    |                                |            |  |                                |            |  |                                |
|------------------------------------|--------------------------------|------------|--|--------------------------------|------------|--|--------------------------------|
| Image Pyramid Schedule             |                                |            |  | 1 1 1 1 1 1                    |            |  |                                |
| Image Sampler                      | Random                         |            |  | Random Coordinate              |            |  | Random Coordinate              |
| Interpolator                       | B Spline Interpolator          |            |  | B Spline Interpolator          |            |  | B Spline Interpolator          |
| Maximum Number Of Iterations       | 1000                           |            |  | 1000                           |            |  | 1000                           |
| Maximum Step Length                | 0.009                          |            |  | 0.015                          |            |  | 0.117188                       |
| Metric                             | Advanced                       | Normalized |  | Advanced                       | Normalized |  | Advanced                       |
| Metric0                            | Correlation                    | Transform  |  | Correlation                    | Transform  |  | Correlation                    |
| Weight                             | Bending Energy                 | Penalty    |  | Bending Energy                 | Penalty    |  | Bending Energy                 |
| Metric1                            | Weight                         |            |  | Weight                         |            |  | Weight                         |
| Weight                             | 0.1                            |            |  | 0.2                            |            |  | 0.1                            |
| Moving Image Dimension             | 3                              |            |  | 3                              |            |  | 3                              |
| Moving Image Pyramid               | Moving Smoothing Image Pyramid |            |  | Moving Smoothing Image Pyramid |            |  | Moving Smoothing Image Pyramid |
| Moving Image Pyramid Schedule      | 0 0 0 0 0 0 0 0 0 0 0 0        |            |  | 1 1 1 1 1 1 0 0 0 0 0 0        |            |  | 1 1 1 1 1 1 1 1 1 1 1 1        |
| Moving Internal Image Pixel Type   | Float                          |            |  | Float                          |            |  | Float                          |
| New Samples Every Iteration        | TRUE                           |            |  | TRUE                           |            |  | TRUE                           |
| Number Of Histogram Bins           | 32                             |            |  | 32                             |            |  | 32                             |
| Number Of Jacobian Measurements    |                                |            |  | 1000                           |            |  | 1000                           |
| Number Of Resolutions              | 2                              |            |  | 4                              |            |  | 4                              |
| Number Of Spatial Samples          | 2048                           |            |  | 2048                           |            |  | 2048                           |
| Optimiser                          | Adaptive                       | Stochastic |  | Adaptive                       | Stochastic |  | Adaptive                       |
| Registration Resample Interpolator | Gradient Descent               |            |  | Gradient Descent               |            |  | Gradient Descent               |
| Resampler                          | Multi Metric                   | Multi      |  | Multi Metric                   | Multi      |  | Multi Metric                   |
| Result Image Format                | Resolution Registration        |            |  | Resolution Registration        |            |  | Resolution Registration        |
| Result Image Pixel Type            | Final B Spline Interpolator    |            |  | Final B Spline Interpolator    |            |  | Final B Spline Interpolator    |
| Transform                          | Default Resampler              |            |  | Default Resampler              |            |  | Default Resampler              |
| Use Direction                      | Tif                            |            |  | Tif                            |            |  | Tif                            |
| Cosines                            | Float                          |            |  | Float                          |            |  | Float                          |
| Write Result Image                 | B Spline Transform             |            |  | B Spline Transform             |            |  | B Spline Transform             |
|                                    | TRUE                           |            |  | TRUE                           |            |  | TRUE                           |
|                                    | FALSE                          |            |  | FALSE                          |            |  | FALSE                          |

**Supplementary table 3.** B-spline transformation parameter configurations for Par00GM, Par0025, Par0033 and Par0038. These parameters are explicitly specified in the parameter map files. Any parameter not specified is set to a default value at Elastix runtime. See Elastix manual for information on default values.

**DEEP LEARNING CELL COUNTING IN ANATOMICALLY TARGETED STRUCTURES**

**ABSTRACT**

A significant challenge of analysing whole mouse brain data comes with the quantification of cell counts. Manual approaches are laborious, highly subjective and exceptionally challenging when data is oversampled and amounts to terabyte-sized data sets. Additionally, identification and delineation of anatomical structures are subjective, presenting the cell counting challenges with multiple compounding sources of error. Here, we took a deep-learning approach to automatically quantify cell counts in oversampled high-resolution serial two-photon data of the whole mouse brain. Registration with the Allen Brain Institute framework atlas allowed for novel targeting of cell counting to anatomical structures of interest, negating the need to process the entire data set. We performed multiple validation steps including the validation of novel augmentation of a small initial data set of both serial-two photon and confocal image data, the validation of the deep-learning detection on novel image data, and the validation of the entire pipeline in the dorsal lateral geniculate nucleus in the Sox14gfp/+ mouse. The pipeline revealed an automated count of  $1,234 \pm 90$  cells (mean  $\pm$  SD) which is comparable to a separate study which found  $1,255 \pm 195$  cells (Evangelio et al., 2018).

## 4.1 INTRODUCTION

Advancement in whole-brain imaging techniques at high resolution enables scrupulous insight into the connectivity and distribution of cell populations across local and distal scales (Y. Kim et al., 2015; Li et al., 2010; Ragan et al., 2012; Renier et al., 2016). These imaging approaches yield gigabyte and terabyte data sets. They can also include multiple imaging channels for multi-fluorophore studies or for acquiring tissue autofluorescence information for use with registration a common coordinate framework (see 'Chapter 3'). However, the analysis of these data sets is very demanding, currently very limited and also highly contingent on the imaging method being utilised. In early studies, quantification of fluorescently labelled data required experts for manual stereological evaluation in individual tissue sections or volumes. Criteria for assessment might include fluorescence intensity, fluorescence density, localisation of fluorescence, the ratio of fluorescence signal to background signal, and morphology either somatic, axonal or dendritic. Of these criteria, basic cell counting provides the most valuable quantification for studies seeking to map and understand cell populations and connectivity. However, similar to the issues surrounding human registration approaches, human analysis of fluorescence imaging data is fraught with issues pertaining to subjectivity, selection bias and overall low throughput owing to time-intensive manual protocols.

In studies utilising cell counting techniques, there are several issues with reproducibility and disagreement between measurements are often reported for the same anatomical regions. Keller et al. (2018) provides an excellent review of total cell counts, neuron counts, astrocyte counts, oligodendrocyte counts, and microglia counts reported in the literature, and it is abundantly clear that independent studies report conflicting numbers. For example, Herculano-Houzel et al. (2013) reported 66,771 neurons in the frontal cortex, whilst Schmid et al. (2013) reported 213,000 neurons; a 1.8-fold disagreement. The dorsal CA1 region of the hippocampus also shows a 2.7-fold disagreement in neuron count between reported numbers of 447,500 (Jinno & Kosaka, 2010) and 165,742 (Miranda et al., 2009). This puts the dorsal CA1 cell count mean at  $306,600 \pm 141,000$  (Mean  $\pm$  SD), with a standard deviation that is comparable to the reported value by Miranda et al. (2009). Reported disparities are simply the result of no standardised process for measuring cell counts in the brain.

Automated stereological counting methods, such as the optical fractionator technique, estimate the total cells number by systematic random sampling of discrete volumes in the structure of interest (Olesen et al., 2017). Such approaches also remove human subjectivity. The limitations of these methods are that structure volumes are discretely sampled and are not comprehensive. For example, an optical fractionator protocol by Olesen et al. (2017) detailing counting of BrdU-positive neurons in the rat hippocampus describes sampling one hemisphere for a specimen and sampling every fifth section at 80  $\mu\text{m}$  section thickness. Volumes are randomly sampled given a manually delineated region for the structure. Sampling volume size is also explicitly chosen to give a combined cell count of 150 to 200 cells per sample, per the structure of interest. The total cell count and density for the hippocampus can be extrapolated using a scaling method and the known structure volume measured from the manual delineation. This counting technique is credited for providing an unbiased three-dimensional approach for cell counting (Olesen et al., 2017), which is insensitive to tissue shrinkage or experimental setup. This is due to strict inclusion and exclusion criteria, and random sampling approaches in the counting protocol. Ultimately, however, this method is far from representing a complete count as a tiny fraction of cells are sampled. It additionally relies on measurable volumes which introduce issues relating to how delineation is defined and also does not address inhomogeneous cell distribution within structures.

Manual quantification techniques on terabyte data sets are a near-impossible challenge to accomplish. A single coronal section image at XY spatial resolution  $0.54 \times 0.54 \mu\text{m}$  acquired through high-resolution STP tomography can be around 400 MB in size using an 8-bit data type (800 MB if using 16-bit data type). At a resolution of  $20,000 \times 15,000$  pixels, an image of this magnitude is hard to view and takes several seconds to load into software such as ImageJ/Fiji. A single whole mouse brain data set could contain a thousand individual images. Achieving accurate quantification of fluorescence signal on a single image is inexplicably challenging, but coupled with oversampling along the axial axis, this requires localisation and tracking of the signal between multiple sections. For example, a single cell could be sampled on four or five concurrent sections if oversampling as per the Nyquist–Shannon sampling theorem. This level of sampling would require a human expert to load terabyte data sets, localise the region of interest, quantify the fluorescence signal such as counting individual cells, and then track these measurements across four or five concurrent sections. All whilst

simultaneously detecting new objects which appear in each section. This is a task which would take several months to accomplish. It also assumes access to a workstation that is capable of handling these high resource demands, and time to be dedicated to such a relentless task.

In light of these limitations, many efforts have gone towards automated methods for cell count quantification. Flow cytometry (FCM) and fluorescence activated cell sorting (FACS) allows fast cell counting and sorting (McKinnon, 2018). This technique requires cells to be dissociated from the tissue and quickly passed through a mechanism to interrogate fluorescence characteristics. However, this technique destroys tissue removing any spatial knowledge of the cell distributions. Volume measurements must come from different in-house experimental studies, or from other sources with different experimental conditions and parameters.

For smaller sample sets such as *Drosophila* brains, automated watershed segmentation algorithms and edge detection approaches have been successfully used for cell counting (G. Lin et al., 2003; Wahlby et al., 2004). At a volume of  $8 \times 10^7 \mu\text{m}^3$  (Zheng et al., 2017), these data sets are a small fraction of the total volume of a mouse brain. Success with threshold-based approaches also relies on a high signal to background autofluorescence ratio. Highly fluorescing molecules such as lipofuscins, elastin and collagen can often appear as intense in fluorescent signal as fluorophore labelled cells (Schnell et al., 1999). The general background signal can also be highly variable attributing to different structures such as grey matter, white matter and vasculature. If working with three-dimensional data sets such as confocal image stacks, laser scattering and artefacts from surfaces of the slice can also impact the overall background signal intensity. Arguably, global thresholding approaches are completely ill-suited for imaging experiments with non-negligible background signal meaning local and computationally complex thresholding approaches are required. However, as background structures in biological tissue can occur over local and global scales, even local thresholding is unfitting.

To this end, we require an approach for unbiased automated detection of cells in three-dimensional data sets such as the whole mouse brain. In recent years, deep-learning (DL) has been gathering significant momentum in classification tasks due to its capacity to extract and

learn features of significance about the data that has been presented to the algorithm. Specifically, convolution neural networks (CNNs) observe improved effectiveness over standard artificial neural networks (ANNs) in image classification and segmentation challenges (Krizhevsky et al., 2017; LeCun et al., 1998). CNNs have consequently seen utility for a litany of biomedical image classification challenges including cell detection on confocal microscopy data sets (Segebarth et al., 2018), epidermal tissue detection (Oskal et al., 2019) and skin cancer classification (Esteva et al., 2017).

Undeniably, it is not hard to find a biological research area where DL models have not applied. Not only do they provide accurate performance either on-par or superior to human performance, but they can also classify large volumes of data in minutes. A mere fraction of the time it might take a human expert. Enticingly, it does not rely on human effort at all. DL models are also unbiased and characteristically perform based on the data that has been presented to the algorithm. Limitations for DL models are therefore contingent on the type and amount of data used in training, as well as a host of hyperparameters that affect the way the models learn and how they generalise to new data sets.

ANNs work by chaining together many nodes (also referred to as units), for which nodes can be distributed across multiple layers. In a feedforward densely connected network, each node in each layer is connected to every node in the adjacent layers. Sparse networks exhibit fewer active connections between nodes of adjacent layers. Each node executes a non-linear activation function on the weighted sum of inputs it receives, plus a bias term. Non-linearity is required to encourage the network to learn non-linear features about data (see below for more information on key components to artificial neural network training). The output of a node is fed forward to the next layer for additional activation. The process repeats until a final output layer is reached and generates a single output value (or values, depending on the application) for the network. The final output can be compared against a known target for classification and regression orientated tasks.

The fundamental component of ANNs is their ability to learn through a back-propagation algorithm. Here the weights for the inputs between nodes are optimised such that the output of the network incrementally improves performance. A CNN is different from an ANN by



introducing convolution steps which seek to extract vital contextual information from dimensionally large data inputs, such as an image. Throwing an image into a fully connected feedforward model would be ineffectual, as each pixel is passed as an input into the network. Any generalised information about the image subject is, therefore, not considered. Images are also information-dense. A small single-channel  $256 \times 256$  image has 65,536 voxels, each of which is an input. Numerous inputs mean increased computational demand on the network. Convolution operations can reduce network complexity by calculating the sum of weighted elements (pixels) using a kernel. The weights of the kernel can change the effect of the operation with some convolution processes leading to feature enhancement operations such as edge detection or blurring. CNNs use convolutions to extract features which become learned by the network, rather than learning from every input voxel.

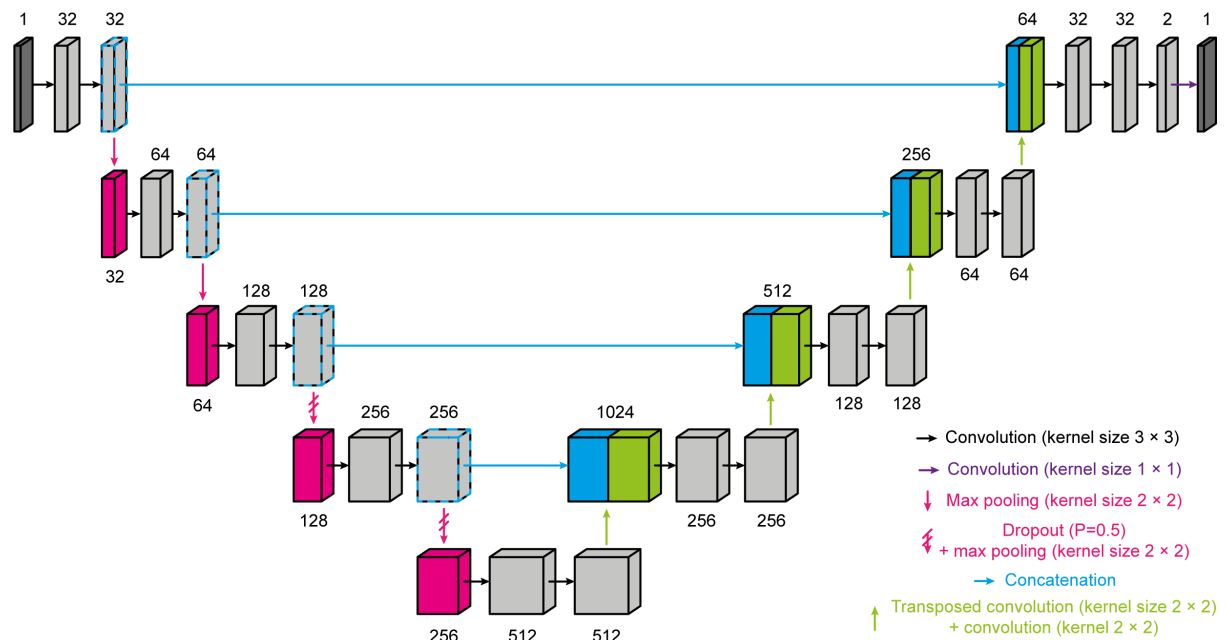
With this in mind, we developed a DL application for automated cell counting, which does not suffer the drawbacks associated with intensity-based segmentation approaches; namely insensitivity to different fluorescence and background signal levels and any signal from non-somatic structures, such as neurites. We used the U-Net architecture (architecture explained below; Ronneberger et al., 2015) to perform binary classification on input images. To make the approach more computationally efficient and faster for large data sets, novel incorporation of registration with the Allen Brain Institute framework allowed counting to be targeted to anatomical regions of interest.

### **Deep-learning segmentation with U-Net**

Cell detection is fundamentally a pixel-wise segmentation task requiring binary classification of pixels into a class for background signal and a class for the signal from fluorescently labelled cells. Of the existing CNN architectures for cell segmentation, the U-Net model proposed by Ronneberger et al. (2015) provides the most ubiquitously used architecture. At its inception, the model outperformed other existing models for which the state-of-the-art segmentation neural network at the time relied on a sliding window attached to a convolution neural network (Cireşan, Meier & Schmidhuber, 2012). The sliding window approach is computationally expensive and also leads to redundant classification where sliding windows overlap. In the biomedical imaging field, imaging data is also relatively sparse. With fewer

examples for classification, it is necessary for learned features to be location invariant rather than spatially constrained.

The U-Net architecture is more focused on scale-invariant and spatially-invariant feature extraction during training. It does this by using two stages in the network (Fig. 4.1). An encoder stage which performs downsampling of input image data through feature extraction, and a decoder stage which performs upsampling of learned key features. Downsampling and upsampling help remove spatial context in the data allowing features of importance to be heavily weighted towards those which are spatially invariant. Successful application for the segmentation of neuronal structures in electron microscopic data and cell detection in phase contrast and diffuse image contrast microscopy images (Ronneberger et al., 2015), has resulted in a surge of adoption of the architecture for other biomedical segmentation challenges. These include liver and lesion segmentation in computer tomography images (Christ et al., 2016), Xenopus kidney structure segmentation (Çiçek et al., 2016) and vascular boundary detection (Merkow et al., 2016).



**Figure 4.1.** U-Net architecture. The network is split into decoder (left) and encoder (right) stages, joined by an intermediate stage (bottom). Numbers represent the number of feature channels at each step with network operations highlighted in the key (bottom-right). Dropout layers are introduced into the last two decoding stages to introduce noise and improve

*network generalisation. Concatenation steps (blue) are required to merge features from the decoder stage with features from the encoder stage of the network. This enhances the important features which are learned by the network during training.*

The U-Net model starts with two consecutive  $3 \times 3$  convolution operations on the input image. The use of paired convolution operations allows extraction of features of increased complexity, over what would be possible with a single convolution operation. For example, a prominent edge in an image of a car would be low level, but the combination of edges could represent corners as being a key feature of the image. A max-pooling process takes the feature maps from the double convolution layer and downsamples the maps by retaining the highest values (and therefore more important features) within a  $2 \times 2$  kernel.

The max-pooling output is fed into another pair of convolution filters followed by another max-pooling process and repeated twice more in total. The successive convolution and pooling steps mean we extract more complex features that remain important to the original image, whilst reducing dimensionality and therefore removing the importance of local spatial information from the input data. After four stages of convolution and pooling, at the lowest level of the encoding stage, there are another two  $3 \times 3$  convolution layers to transition the network into the decoding stage.

The goal of the decoding stage is to upsample the image back to the original input size using the key features extracted during encoding. Upsampling is initiated using a  $2 \times 2$  transposed convolution process. The transposed convolution process uses a transposed convolution matrix meaning the matrix multiplication increases dimensionality in the output (Zeiler et al., 2010). Thus, the encoded feature maps at the bottom of the network can be upscaled. To increase the accuracy of the network, feature maps from the corresponding network in the encoding path are extracted and concatenated with the upsampled feature maps. This step is known as a skip connection. Supplementing the upsampled features with feature information from the encoding stage in this way, means important information that resides in the encoding feature maps can also be used to reinforce key features for detection. For the key features across the encoder and decoder stage to be extracted, the skip connection is followed by an additional double convolution step.

The transposed convolution, skip connection concatenation, and double convolution process is repeated three times. At the final stage, a trio of convolution processes is required to get an output image with the same dimension as the input.

During training, the network is fed with image data and ground-truth segmentation data. As image data is passed through the network, the output segmentation results are compared to the ground-truth and weights distributed throughout the network are modified, such as to encourage the network to generate increasingly accurate predictions. There are a handful of key components for training the U-Net, and any neural network in general. These are described below.

### **Activation function**

The activation function computes a weighted sum of inputs, with an additional bias term which adds a tuneable constant that is independent of the inputs. The output of this function is known as the activation for which the magnitude of the activation is analogous to declaring whether the node is active (or involved in the network computation). The key property for activation functions is that they are non-linear which has a few benefits for a network. First, non-linear activation functions reflect non-linearity in real data sets. Second, in order for the network to learn, the activation function must be differentiable and complex prediction challenges with multiple layers of activation functions, the gradient can still be calculated across all the layers with respect to the inputs. This is vital for network learning, as a gradient descent optimisation function (discussed further below) during the back-propagation process (learning phase) is able to calculate how wrong an initial prediction is with respect to the node inputs, and how much to modify the weights in the activation functions. A linear function is ineffective as its gradient is independent of the input and constant. The gradient would, therefore, not indicate by how much the weights in the activation function should be modified, other than to say that the weights should be modified. Third, linear functions are unbound, meaning the gradient does not push the activation function towards a stable value at either end of the function.

Commonly used activation functions in ANNs include Sigmoid (Han & Moraga, 1995), Tanh, ReLu (Nair & Hinton, 2010) and ELU (Clevert et al., 2015). ReLu and ELU provide the most favourable activation functions for complex networks such as the U-Net. This is because both ReLu and ELU derivatives are mathematically simple to derive, plus the activation (output) of ReLu and ELU can be zero or negative meaning a densely connected network can have nodes which have zero transfer to the next layer. This reduces the density and complexity of the network itself by forcing unimportant nodes into an inactive state.

### **Loss function**

Network learning is formulated as an optimisation challenge on the weights in a network, by using a known ground-truth data set for network error evaluation. For the network to know which weights to modify, and by how much to modify them, a loss function is used to measure network accuracy given the current state of learned predictions. Loss functions are more complex than a simple linear accuracy measurement as the network needs to learn how wrong it is in order to maximise learning. For example, an initial network accuracy of only 10% would see the same encouragement of weight modification during back-propagation at 20% or even 90% if the loss function is linear. If using a log loss function such as cross-entropy, the loss value at 10% accuracy would be much higher than that at 90%. Consequently, at a poor network performance accuracy of 10%, the high loss value forces the network to alter the weights to drastically improve performance. When the network is performing desirably at 90% accuracy, a smaller loss value restricts drastic weight modification. If the weights change too much, the network loses the ability to refine its learning and becomes more stochastic in its performance. For a binary image segmentation network, binary cross-entropy loss offers the most favourably used loss function. The Dice score which measures similarity between binary images can be reworked as a loss function, and the same is true of the intersection-over-union score (also known as Jaccard score). Recently, focal Tversky loss appeared as a modified Dice loss score which has a  $\alpha$  parameter that can be used to focus the loss towards penalising false-positives ( $\alpha < 0.5$ ) and false-negatives ( $\alpha > 0.5$ ) (Abraham & Khan, 2018). This metric is useful for encouraging the network to minimise false-rate occurrences which has utility in biomedical image segmentation applications.

### **Optimiser**

An optimiser works closely with the loss function to understand how to optimise the weights in a network during back-propagation. Optimisers are gradient descent algorithms which aim to find a minimum in the activation function, which reflects an optimum weight configuration. This is why the activation function must be differentiable. Restrictions for learning can be implemented during the back-propagation process with a learning rate parameter. This parameter is  $<1$  and scales down the gradient to prevent overly broad weight modifications which could result in the network finding local minima instead of global minima. In practice, a learning rate is always implemented and typically with a value  $<0.1$  to ensure network learning is stable.

One optimiser method is stochastic gradient descent (SGD) which uses an approximation of the network gradient by only evaluating a subset of samples to improve learning efficiency and is commonly used. Adaptive moment estimation (Adam) is another gradient descent approach which uses prior gradient estimations to give momentum for the optimiser to help find the minima more efficiently.

## 4.2 METHODS

Here we use the U-Net architecture in a pipeline to automatically segment fluorescently labelled STP image data using the Allen CCF3 registration result (see 'Chapter 3') to target cell count quantification to particular anatomical regions of interest. To this end, we used several optimisation steps to generate the optimum U-Net network. This includes the generation and characterisation of manually segmented data, evaluating data augmentation processes for use in supplementing a small initial data set, choosing the optimum loss function to get a low validation loss (best performance) during training, tuning of additional hyperparameters including optimiser and learning rate, and the characterisation of the trained model itself. Lastly, we validated the complete pipeline with automated counting of GFP+ cells in the thalamus of the Sox14gfp/+ mouse. All pipeline and analyses were done with Python, using Keras (Chollet & others, 2015) and TensorFlow 2 (Abadi et al., 2015) for all machine learning implementation.

### **Sample preparation for STP imaging and pipeline validation**

Sox14gfp/+ brains for STP imaging were extracted following a transcardial perfusion process using 20ml 1X phosphate-buffered saline (PBS) at 4°C, followed by 20ml 4% paraformaldehyde (PFA) at 4°C. Brain tissues were post-fixed in 4% PFA and kept at 4°C for 24 hours before transferring to 1X PBS with 0.01% NaN<sub>3</sub>. To embed the sample in preparation of imaging, a 4.5% oxidised-agarose solution was prepared by stirring agarose (type 1; Sigma) with 10 mM NaIO<sub>4</sub> (Sigma) and 50 mM phosphate buffer (PB) at room temperature in a light-shielded environment for 2 hr 30 mins. The oxidised-agarose solution was filtered three times with PB before re-suspending in PB and stored at 4°C until performing the embedding process. During embedding, two 25G needles were horizontally inserted ~5 mm apart, through a disposable cube-shaped embedding mould (VWR #E6032-1CS) to provide a horizontal platform to support the mouse brain. The oxidised-agarose solution was heated to boiling in a microwave and left to cool until 65°C. After filling the mould with the agarose solution, the mouse brain placed onto the platform using a set of forceps. The needle platform helped support and align the sample and meant the entire intact sample could be imaged without necessitating slicing a flat surface onto the tissue. Finally, the embedded sample was left to passively cool to room temperature where it became solid, before carefully extracting

from the moulds and immersed and stored within 1X PBS with 0.01% NaN<sub>3</sub> at 4°C until imaging was performed.

### **STP imaging for pipeline validation**

Agarose embedded samples were glued with a cyanoacrylate-adhesive (Loctite Ltd.) onto a glass slide, with the Anterior-Posterior axis of the mouse brain was positioned orthogonal to the slide. The base of the slide consists of two magnetic blocks which allowed the sample to be magnetically fastened to a magnetic base in a custom-designed 1.5L plexiglass water-bath. A custom-designed cooling system (JULABO UK Ltd.) was inserted into the bath and set to maintain a 4°C environment to maintain tissue integrity over several days of imaging. The bath was then filled with PB and translated to align the sample with the imaging objective and microtome. For image acquisition, physical sections were acquired every 50 µm using the microtome, with five optical sections using a piezo controlled objective acquired per physical section, providing a 10 µm axial resolution. These were combined with a 16x, 0.8NA imaging objective (Nikon Inc.) to obtain a spatial resolution of 0.54 µm. The first optical layer was acquired 25 µm below the surface to avoid surface distortions. Complete coronal sections of the mouse brains were acquired by tiling the objective 1x1 mm field-of-view across the coronal plane. This resulted in a 12 x 10 tile mosaic for each optical section. Physical sections were obtained using a microtome with a blade vibration frequency of 60 Hz and velocity 0.35 mm/s. Laser (Chameleon Ultra II, Coherent) excitation was conducted at 920 nm for GFP excitation. The laser power was incrementally increased for optical section acquisition at deeper depths to compensate for laser power loss induced from scattering. Emission fluorescence was collected over three photomultiplier tubes (PMTs) channels (red, green and blue) with the green channel path consisting of an additional 525/40 nm bandpass filter (Laser 2000 Ltd.) to reduce collection from non-GFP sources. PMT gain was set to 850V for all three channels. Tile acquisition was conducted with a resolution of 2080 x 2080 pixels with 16-bit range, amounting to approximately 3 megabytes per tile. Imaging conducted with these parameters resulted in 250 optical sections along the Z-axis (caudal-rostral axis) with each section comprising 120 tiles. The acquisition occurred over ~5 days and generated ~3.5 terabytes of data per brain. All images were stored on a network-attached storage (NAS) drive (DS1815+; Synology Inc.). Following imaging, free-floating slices were extracted from the PB and stored in 1X PBS with 0.01% NaN<sub>3</sub> at 4°C.



### **Image stitching pipeline**

Individual tiles acquired through STP imaging were stitched together with a custom pipeline written in Python and executed alongside STP image acquisition. Once all tiles per physical section were obtained, each tile was cropped to 2024 x 2024 pixels to remove artefacts on the edges of each tile. A previously acquired average flatfield image was gaussian blurred with kernel size 201 x 201 pixels and sigma = 30. This was done to remove high-frequency structures in the flatfield image and to obtain a low-frequency background intensity profile. The flatfield image was then divided by its maximum intensity value and subtracted away from integer value 2. This converted the flatfield image into a correction image. An intensity ramp defined by a quarter-circle function was applied to a 144-pixel wide boundary along each edge of the correction image resulting in a non-linear reduction in pixel intensity to zero at the absolute edge of each tile. The 144-pixel wide boundary was chosen as this was the remaining pixel overlap between neighbouring tiles. Intensity ramping was done to reduce tiling artefacts when eventually stitching the tile images together. Each STP tile was multiplied by the intensity-ramped correction tile and stitched using the Grid/Collection stitching plugin in ImageJ/Fiji using maximum intensity projection along the overlapping regions. Stitched images were further converted to an 8-bit range by appropriately scaling the 16-bit image range, and additionally downsized to an isotropic voxel size of 10  $\mu\text{m}$ .

### **Registration using Allen Institute Common Coordinate Framework (CCF3)**

For delineation of the mouse brain into anatomical structures, SimpleElastix (Klein et al., 2009) was used to perform a rigid, affine and b-spline transformation of the CCF3 atlas to the stitched STP data. First, stitched STP data was downsized to a 10  $\mu\text{m}$  isotropic voxel size. Parameter configuration follows Par00GM (see 'Chapter 3'). A multi-step multi-resolution registration of the average CCF3 atlas to the STP data was performed beginning with a single-resolution rigid and affine transformation using an advanced normalized correlation metric optimised through an adaptive stochastic gradient descent algorithm. This was followed by a four-resolution b-spline transformation using the Mattes mutual information metric and the adaptive stochastic gradient descent optimiser, with a final grid spacing of 250  $\mu\text{m}$  over 1000 iterations. The transformation map was used to warp the annotation CCF3 atlas and a custom

hemisphere atlas (which delineates the two hemispheres in the CCF3 atlases) with a nearest neighbour interpolation.

### **Manually segmented training data**

Training data comprised an initial 151 images ( $512 \times 512$  pixels;  $276 \times 276 \mu\text{m}$ ) randomly sampled from exemplar STP data sets containing GFP fluorescence labelling localised to soma, dendrite and axonal structures (Fig. 4.2A). Each training image was loaded into ImageJ/Fiji and manually segmented with the brush tool. Criteria for cell segmentation were as follows. Cells chosen for masking had to exhibit sufficient signal over the background to delineate the shape of the soma, and to constitute a cell-like signal. The cell mask had to follow the morphology of the soma as accurately as possible. Any fluorescence from dendritic or axonal structures was omitted from masking. In the case where cells were closely positioned and appeared to overlap, but could be visually distinguished as separate cells, the cell masks were made to explicitly separate the two objects with a one- or two-pixel separation. We chose to do this to encourage the network to learn features that distinguish individual cells when they appeared in densely packed clusters. In cases where cells were too densely arranged so individual cells could not be visually discerned, no effort was made to distinguish between cells in the mask image. If a human was incapable of distinguishing these structures, then it was deemed an unrealistic expectation for a neural network to do so. Therefore, this was not considered detrimental to the performance of the trained network.

The density of the cell fluorescence class ranged from zero to around 10% of the image with a mean of 1.23% ( $n=151$  images). Images containing zero cell fluorescence were kept in the data set to provide examples of background signal to the U-Net during training. These images typically showed autofluorescence, vasculature structures and artefacts exhibiting high or varied fluorescent signal such as the meninges membranes. Each pair of STP image and the manually segmented image was saved as a single-channel 8-bit TIFF.

In order to improve generalisability performance of the network and to supplement the training data with more examples, an additional 17 images ( $1024 \times 1024$  pixels) were obtained from <https://github.com/matjesg/DeepFLaSH> (Segebarth et al., 2018). These images were already used for a DL approach for confocal image cell segmentation and were considered

appropriate for use here. Each image already had a manual segmentation mask. However, the manual labelling of this data set was done by using elliptical markers placed over each cell rather than using the criteria outlined earlier. As the goal was to improve generalisability of the network, this slight incongruity in segmentation protocol was not considered detrimental to the network performance. The original 151 training set, referred to as GM data, has training samples of size  $512 \times 512$  pixels. The supplemental data from Segebarth et al. (2018), referred to as MG data, was therefore split into four blocks of  $512 \times 512$  pixels to create a total of 68 images.

We characterised the GM data ( $n=151$ ), MG data ( $n=68$ ) and combined GM and MG ( $n=219$ ) for class balance between cell and background, by measuring the number pixels corresponding to a cell classification.

### **Augmentation of training data**

For biomedical imaging data, a combined GM and MG data set ( $n=219$  samples) was considered a reasonably sized sample set. However, to improve overall performance and to encourage generalisation by the network, a series of augmentation steps was performed to bolster the number of training samples by several folds. Poisson and Gaussian noise were artificially added to all data sets (GM, MG, and GM and MG combined) to encourage the network to learn features which were noise invariant. Measuring image noise was a challenging task in itself. To assess how Poisson and Gaussian noise addition influence the noise in the data sets, image noise variance was calculated using the method proposed by Immerkær (1996) which uses the following  $3 \times 3$  kernel,

$$k = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \quad (\text{Eq. 4.1})$$

convolved and summed across each image. This approach had the benefit of being fast and gave a reasonable estimation of noise variance. Evaluation of noise added to the data (see ‘Discussion’) resulted in Poisson noise being chosen as the only noise addition to be performed on the data set during augmentation.

With this in mind, the combined GM and MG data set was augmented in the following ways. First, vertical flipping, horizontal flipping,  $90^\circ$  rotation,  $180^\circ$  rotations and  $270^\circ$  rotations were

applied to all 219 images and masks pairs. Combinations of flipping and rotations which resulted in the same result were omitted, so no augmented sample image was identical. In general, the initial flipping and rotation transformation is recommended to help the network learn rotation- and location-invariant features. Stage one augmentation step resulted in an 8-fold increase in data to  $n=1,752$  samples from the combined  $n=219$  data set.

We implemented a second augmentation stage using the Augmentor Python library (Bloice et al., 2017) to perform the following operations; rotating the images in the range  $-360^\circ$  to  $+360^\circ$ , zooming in the range 90% to 110%, skewing along the horizontal or vertical axis, adding a random elastic distortion using a grid size of 10-by-10 pixels, shearing along the horizontal or vertical axis, and adding Poisson noise. All operations except the Poisson noise addition utilised native Augmentor operations. Poisson noise addition (chosen over Gaussian noise) was achieved by creating a custom class which interfaced with the Augmentor operation pipeline. Each operation had a 50% probability of occurring for every image passed through the pipeline. To ensure image data and mask data were augmented in the same way, we fed an image and mask pair into the pipeline with the same probabilities. Each generated image and mask pair were saved into the same directory as the input data. The number of images to generate was specified before training. In this manner, the U-Net was dynamically fed with supplemental data alongside the initial stage one augmentation data set. Care was taken to limit the magnitude of the operations to prevent unrealistic augmentations.

During network training, a 6-fold augmentation at stage two generated 12,264 training samples which were split 75% (9,198 samples) for the training data and 25% (3,066 samples) for the validation data. As augmentation operations were performed randomly, it was a possibility that splitting the data would result in a different balance between cell class and background class across the training and validation sets. All images and masks generated through the augmentation process were therefore stored with a filename referencing the original input image filename. The native filesystem ordering arranged the input data such that all augmented images are immediately listed after the original image, which conveniently resulted in grouping together of images with similar data balance. We then split the data into training and validation by sampling every fourth image for the validation set (25%) with the remaining images used for the training set (75%). Using the native filesystem ordering for

sample splitting ensured the training and validation data sets retained the same class balance and prevents the network generating loss values on the training data which were unrepresentative of the balance in the validation data.

### **Loss function evaluation**

To optimise the segmentation network loss function, we created three simulated faux prediction data sets. Each data set was designed to reflect loss function handling concerning true-positive, true-negative, false-positive and false-negative occurrences. All simulated prediction data sets featured a ground-truth square object labelled with value 1 against a background class of value 0, therefore simulating a binary classification problem. Each prediction set featured a total of eight predictions against the ground-truth, starting with an inaccurate prediction and improving accuracy with each subsequent faux-prediction as measured using the intersection-over-union score (IoU; also known as Jaccard score). Evaluated loss functions included binary cross-entropy loss (BCE), distance map loss (Caliva et al., 2019), binary focal loss (T. Y. Lin et al., 2020), Dice loss, focal Tversky loss (FTL) and combo loss (Taghanaki et al., 2018). BCE loss presents the most ubiquitously used loss metric for binary segmentation. Focal Tversky loss recently appeared as a loss metric for segmentation networks which is similar to the Dice loss but enables weighting the training towards minimising false-negatives or false-positives occurrences.

Simulated prediction set 1 (translation to match) featured a predicted square object identically sized to the ground-truth but started with an initial prediction to the left of ground-truth prediction which progressively translates to the right until total incidence over the ground-truth object (Fig. 4.3). This simulated data set exhibited true-positives, true-negatives, false-positives and false-negatives for which false-positives and false-negatives were equal in occurrence for every simulated prediction.

Prediction set 2 (scale small-to-match) featured an initial square prediction which was spatially centred over the ground-truth, but smaller in scale. Over simulated predictions, the predicted object incrementally scaled until complete incidence with the ground-truth (Fig. 4.3). This simulated data set exhibited true-positives, true-negatives and false-negatives only.

Prediction set 3 (scale large-to-small) was similar to prediction set 2. However, the initial prediction was much larger than the ground-truth and incrementally scaled smaller until complete incidence with the ground-truth (Fig. 4.3). This simulated data set exhibited true-positives, true-negatives and false-positives only.

For each prediction data set, the occurrence of true-positives, true negatives, false positives and false negatives was measured. We then calculate a loss value for each simulated prediction data set for each loss function. For the FTL loss which could be modified to target the loss towards false-positives or false-negatives through an additional parameter,  $\alpha$ , an additional test was done which altered  $\alpha$  from 0.1 to 0.9 where  $\alpha < 0.5$  targets false-positive minimisation and  $\alpha > 0.5$  targets false-negative minimisation.

### **Hyperparameter tuning**

To optimise hyperparameter values for U-Net training, we conducted several small-scale test U-Net training experiments. In the first test, the ELU and ReLU activation functions were compared. In the second test, the loss function was compared between BCE and FTL with FTL  $\alpha$  additionally modified from 0.2 to 0.8 in increments of 0.2, for progressive weighting between false-positive learning and false-negative learning. In the third test, the optimiser was varied between Adam and SGD with learning rates set at  $1e-3$ ,  $1e-4$  and  $1e-5$ . In the final test, the amount of stage two augmentation was varied between 0, 4, 6 and 8-fold increases. In all but the last test, augmentation was limited to stage one only (flipping and  $90^\circ$ ,  $180^\circ$  and  $270^\circ$  rotations). Each test was conducted with early stopping evoked after 30 epochs of no loss improvement to restrict computation time and resources. Unless the specific parameter was modified as described above, the base parameter configuration of the model during all training scenarios was using the ELU activation function, a BCE loss, an Adam optimiser initialised with a learning rate of  $1e-4$ , a batch size of 8 to feed the network with training samples, and Dice loss score for assessing the best model. All test scenarios were run on an Ubuntu Intel(R) Core(TM) i9-7980XE CPU at 2.60GHz with 32 cores and 128 GB RAM workstation equipped with an NVIDIA Titan Xp GPU.

### **Optimised U-Net model training and evaluation**

The final optimised U-Net model was trained using the ELU activation function with an Adam optimiser operating on the BCE loss function. We used a batch size of 8 with early stopping evoked if the validation Dice loss did not improve after 30 epochs (iterations of training). Only the weights which gave the best validation loss during training were saved, rather than saving the model at the last training epoch. Alongside stage one augmentation, stage two augmentation was set to augment 6-fold giving a total 12,264 training samples which were split 75% (9,198 samples) for training data and 25% (3,066 samples) for validation data. The network training was done on an Ubuntu Intel(R) Core(TM) i9-7980XE CPU at 2.60GHz with 32 cores and 128 GB RAM workstation equipped with an NVIDIA Titan Xp GPU. After training, the U-Net model was able to be set up for inference meaning the model was used for prediction only and not for training. In this state, the model could receive an input image and generate a prediction map based on what the model had learned in training. The prediction map had a one-to-one pixel correspondence with the input image. However, pixel values were probabilities in range 0-to-1 representing confidence towards the cell class (value of 1) or background class (value of 0). Therefore, a threshold at 0.5 was taken, to convert the prediction map into a binary mask.

Often, network performance is quoted using the final validation loss, or by using the model to predict the training set or a subset of the training set, and evaluating a number of metrics such as precision, recall and F1 score. Arguably, these reported performance scores only reflect the model performance as a state of training and do not reflect deployment performance when the model is exposed to novel data. We performed model evaluation using a simulating a real deployment condition by using the trained U-Net to classify two novel coronal slices extracted from two STP data sets which were not used to generate the training data. Choosing novel images ensured the model had not been exposed to image data during training which was indirectly related to the novel images tested during validation. Each novel image (approximately  $20,000 \times 15,000$  pixels) was manually segmented to generate a ground-truth mask. Novel image 1 featured soma localised GFP fluorescence distributed across the thalamus. Novel image 2 featured soma, dendritic and axonal GFP fluorescence localised to the dorsal lateral geniculate nucleus (LGd). Each image and its corresponding ground-truth mask was split into  $512 \times 512$  pixel blocks and classified with the U-Net model. Splitting the novel images into blocks was done to utilise the GPU for prediction, for which

the image size would otherwise be too large to predict. Predicted results for each block were used to reconstruct the image into a corresponding image mask and thresholded at 0.5 to generate a binary mask.

Prediction results were compared against the ground-truth by evaluating the following metrics; the IoU, pixel accuracy, precision, recall and F1 score. Evaluating a model based on these metrics is often not too insightful to performance as segmentation can vary widely from the ground-truth without necessarily performing poorly. For example, a cohort of subjects could be tasked with segmenting cars in an image of a car park. Everyone would be successful at identifying the cars, but some people might opt to include car wheels into the mask, and others might choose to omit them. Metrics such as precision, recall and F1 score are sensitive to inaccuracies between masks on an individual pixel basis and would reflect this in their scores. Nevertheless, no-one would deny that any performer was poor at classifying a car. To factor this into evaluating the model, we created a colocalisation metric which acts as a hit-rate for detected cells and ignores the exact segmentation as it is not pixel-dependant. The metric was calculated as the ratio of the number of predicted objects which hit (colocalise), over the number of ground-truth objects. A predicted object was defined as hitting if the segmented object colocalised with the central position of a ground-truth object.

For additional evaluation of the model, both novel images were presented to an independent human performer (not the same performer who generated the ground-truth) to segment both images manually. Both manual predictions were compared against the ground-truth using the same metrics as detailed above.

### **Anatomically targeted cell counting using trained U-Net**

The method presented here covers the complete analysis pipeline using the registration result alongside the trained U-Net model. Automated cell detection on novel STP data sets begins by searching the registered annotation result for anatomical regions of interest. Regions were identified by using the ABA hierarchically organized taxonomy which identifies structures in the annotation atlas given their pixel value. For example, the LGd has a pixel identification value of 170, meaning pixel values of 170 in the annotation atlas correspond to the LGd. An important observation of the annotation atlas and corresponding taxonomy is that not all



structures listed in the taxonomy were located in the annotation atlas. Many structures are nested into substructures for which some may not be present in the atlas. For example, the earlier example of the LGd, with an identification value of 170, does not exist in the atlas but is comprised of the core (LGd-co), shell (LGd-sh) and ipsilateral zone (LGd-ip) for which only the LGd-sh exists in the atlas with an identifier value of 496,345,664. Chosen structures were, therefore, used to locate all substructures which do exist before processing the data.

Each substructure was extracted as a mask from the registered annotation atlas, and corresponding registered hemisphere atlas, before upscaling to the original  $0.54 \times 0.54 \mu\text{m}$  voxel size using nearest-neighbour interpolation (this prevented pixel value interpolation and corruption of the mask at the boundaries). Large nearest-neighbour upscaling resulted in a pixel-like boundary corruption, so a large median filter (kernel size  $k=120$ ) was chosen to smooth the structure into a more anatomically accurate mask before being applied to the original STP image data. The kernel size for the median filter was chosen by using the LGd boundary and applying different kernel size dimensions until there was an optimum visual correspondence between the median-filtered boundary and the true anatomical boundary. The computational scaling law for the median filter process means larger median filter kernel sizes took much longer to execute with not observable improvement to the delineation correspondence.

To reduce memory consumption (owing to the analysis of terabyte data sets), an accompanying coordinate map was generated which maps pixel location by original coordinate position in the STP data. This allowed cropping and removing unnecessary blank data without losing spatial contextual information about each pixel that remained. The cropped image, coordinate map and hemisphere map get passed to the U-Net segmentation model for prediction and thresholded at 0.5 to binarize the result.

To improve the throughput of model inference and to circumvent non-native parallelisation of Keras and TensorFlow, the model was hosted as a Flask web server allowing different threads to handle multiple predictions. Additionally, the U-Net featured four successive downsampling stages meaning the input image sizes had to be evenly divisible by 16 or cause an error. Images required padding to the nearest multiplicative size. This did not affect spatial

localisation as padding was done on the trailing right and bottom edge of each image, whereas the matrix coordinate origin is positioned at the top-left.

Detected cell locations were saved as a list of coordinates locations to file with an additional key for their hemisphere location. Once all relevant images containing the structures for analysis were processed, the list of detected cell positions was corrected for oversampling in the Z-axis by grouping positions which colocalise within a defined radius of 10  $\mu\text{m}$  and only retaining the cell coordinate at the central Z-slice position. The final output from the pipeline was a separate file for each structure which features a list of coordinate positions and hemisphere residence, where coordinate positions are corrected for oversampling if required.

## 4.3 RESULTS

### Manually segmented training data

ImageJ/Fiji provided a practical approach to manually segment data for network training (Fig. 4.2A). Whilst training samples included fluorescent signals from the soma and neurites, the goal of the DL network was to detect cells based on features of the soma only. All other fluorescent structures, such as dendritic and axonal signals, were omitted. Manual segmentation of the GM samples ( $n=151$ ) took over a week to complete, illustrating the time-intensive procedure for manual analysis of data sets. For 151 samples at  $512 \times 512$  pixel size, this corresponds to a manual classification of 40 million pixels. The MG confocal sample set already contained segmentation masks for each image. However, these masks were generated by placing an elliptical mask over each cell position and therefore was not an accurate delineation of the soma, but rather an approximation. Ideally, both data sets should be curated in the same manner. However, as one of the goals here was to provide a generalised network, the difference in masking protocol between the two data sets was not deemed detrimental to the network performance. It would potentially encourage the network to learn more general features for cell detection that were not biased by a single masking protocol.

Both the GM and MG data set displayed substantial class imbalance with the combined data set used for training exhibiting 1.44% cell class labelling and 98.56% background class labelling ( $n=219$ )(Fig. 4.2B). Furthermore, a histogram of the class distributions shows the vast majority of the samples displayed  $<1\%$  cell class balance with image samples either very sparsely populated with cells or completely devoid of cell fluorescence and displayed complete background signals instead.

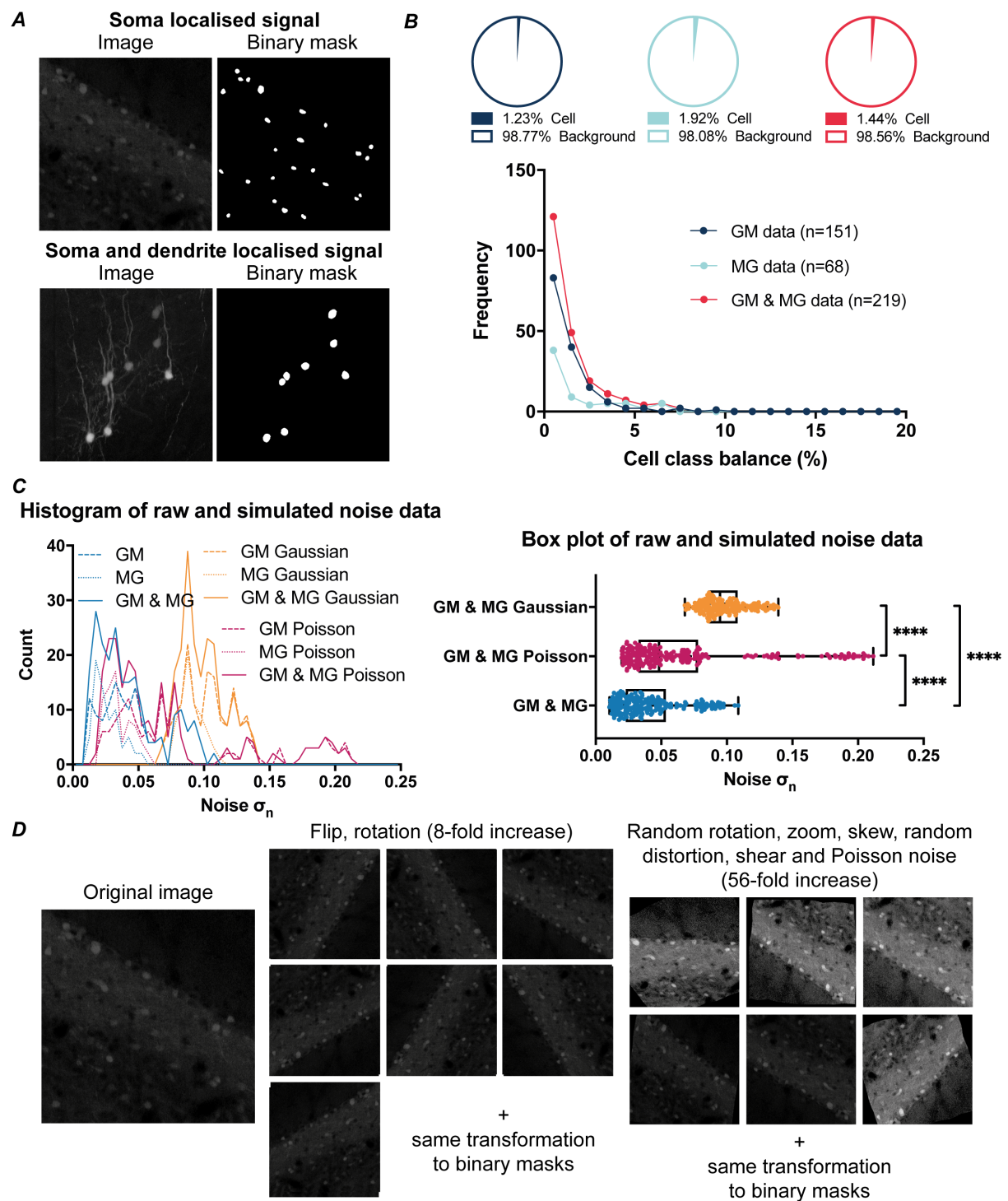
### Augmentation of training data

Augmentation of the training data permitted a remarked increase in sample size without requiring additional data to be manually segmented. To make the network more general and noise-invariant, additional noise was added to the sample data through a Gaussian or Poisson process and then measured using the noise estimation approach by Immerkær (1996). The original GM and MG data set showed a Poisson binomial distribution of noise (Fig. 4.2C) between 0 and  $0.1\sigma_n$ , where  $\sigma_n$  represents the standard deviation of the noise (Immerkær,

1996). Addition of Gaussian and Poisson noise significantly increased and skewed the image noise level in comparison to the original combined GM and MG data set ( $p < 0.0001$ ,  $n = 219$  samples). Significance was tested using multiple comparisons test with one-way ANOVA using matched rows, assuming non-equal variability of differences with a Geisser-Greenhouse correction. Adding Poisson noise to the data increased the range of the noise to  $0.2\sigma_n$ . In contrast, the Gaussian simulated noise reduced the range of noise but shifted the median from an original value of around  $0.03\sigma_n$  to  $0.1\sigma_n$ . Combining both Poisson and Gaussian noise simulation on data generated completely unrealistic training samples as the noise level was too high and in most cases, completely obscured fluorescence signal from the cell. As Poisson noise increased the level of noise in the training samples in a realistic manner, Poisson noise was kept for noise augmentation whilst Gaussian noise simulation was omitted.

The augmentation pipeline contained two stages. Stage one augmentation included horizontal and vertical flipping combined with  $90^\circ$ ,  $180^\circ$  and  $270^\circ$  rotations. Applying all combinations of transformations generated twelve images, for which four have duplicate results. Duplicate images were omitted to prevent duplicated training images which could lead to overfitting (Fig. 4.2D). Stage one resulted in an 8-fold increase in training data which included the original 219 samples. Stage two augmentation used the Augmentor package to provide additional transformations in the form of rotating the images in the range  $-360^\circ$  to  $+360^\circ$ , zooming in the range 90% to 110%, skewing along the horizontal or vertical axis, adding a random elastic distortion using a grid size of 10-by-10 pixels, shearing along the horizontal or vertical axis, and adding Poisson noise. A pipeline executed each operation with a 50% probability of occurring, with additional random probabilities associated with the magnitudes of each operation. Consequently, the same operation invoked for two different input image and mask pair, did not result in the same transformation. Each augmented image and mask pair were, therefore, unique. For a six-fold augmentation factor with the pipeline (every input image and mask pair were augmented six unique ways), this resulted in a 56-fold increase from the original training data set when combined with the initial stage one augmentation. Data augmentation was, therefore, a very powerful and useful approach to bolstering training data, particularly when working with biomedical imaging data where access to original data is sparse with high variability between biological subjects. Despite very high augmentation on a low number of raw image and mask pairs, augmentation benefitted most from the random

elastic distortion and noise, which changed the distribution of pixel values with noise or interpolated values.



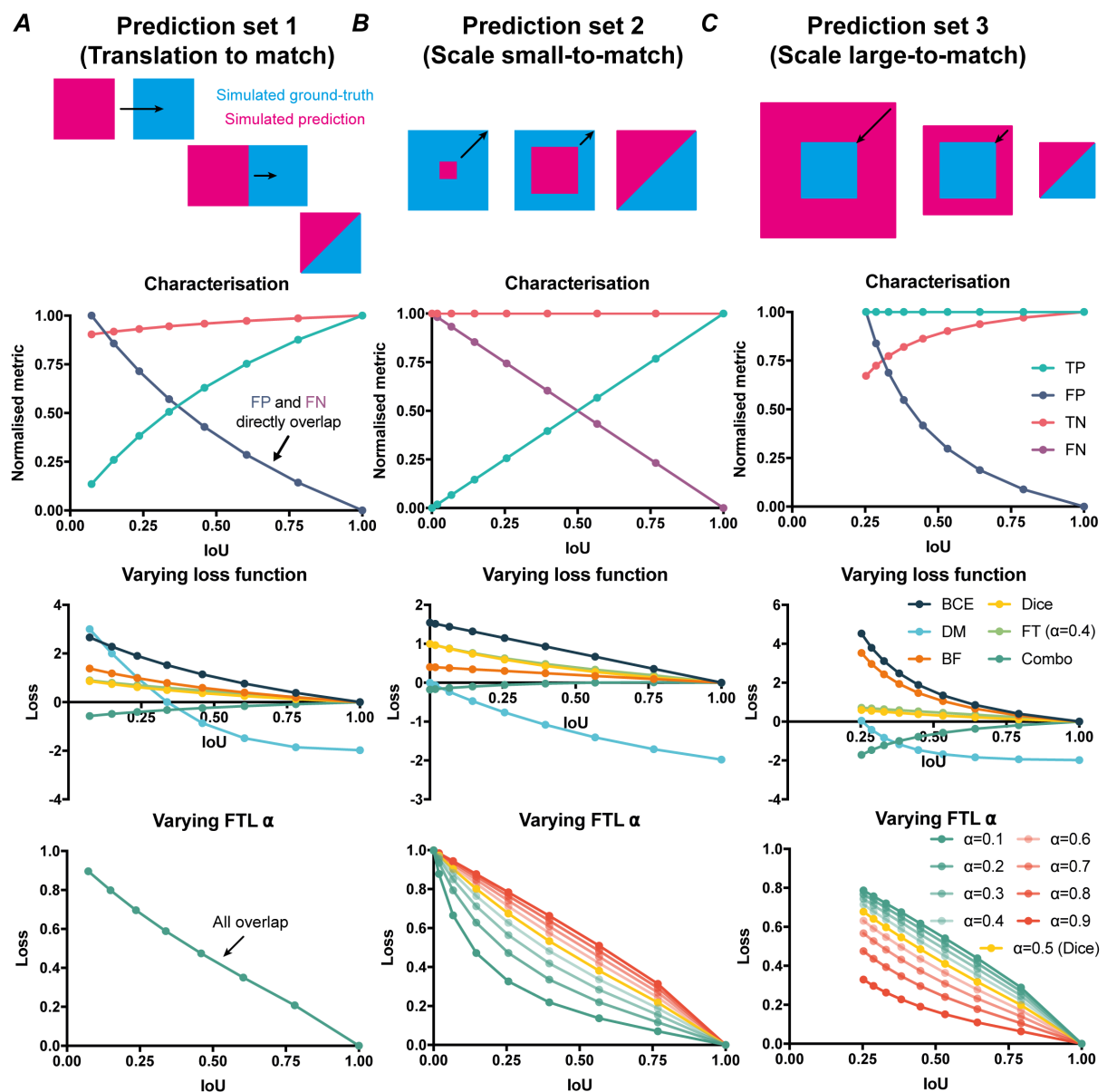
**Figure 4.2.** Curation of training data for the U-Net. **A.** Training images ( $512 \times 512$  pixels) are manually segmented for cell class (value 1) and background class (value 0) to generate binary masks. **B.** Characterisation of the training data sets for GM samples and MG samples (Segebarth et al., 2018). There is a significant imbalance between both classes with the

combined data set ( $n=219$ ) having only 1.44% cell class and 98.56% background class. The majority of images contain <1% cell class balance. **C.** Histogram and box plots of estimated noise standard deviation ( $\sigma_n$ ) following Poisson and Gaussian noise simulation in training data sets. Both Gaussian and Poisson noise addition algorithms increased image noise, with Poisson noise increasing the range of noise levels up to  $0.2\sigma_n$ . Box plots show minimum, first quartile, median, third quartile, and maximum noise levels. Both noise simulations generate noise distributions which were significantly different  $p<0.0001$  ( $n=219$ ) tested with a multiple comparisons test with one-way ANOVA using matched rows, assuming non-equal variability of differences with a Geisser-Greenhouse correction. **D.** Image augmentation pipeline executed in two stages. In stage one, original images were horizontally flipped, vertically flipped and rotated by  $90^\circ$ ,  $180^\circ$  and  $270^\circ$  rotations. Duplicate images after combining transformations are omitted, resulting in an 8-fold increase in training data. Stage two augmentations are implemented with Augmentor. Images are transformed by rotating the images in the range  $-360^\circ$  to  $+360^\circ$ , zooming in the range 90% to 110%, skewing along the horizontal or vertical axis, adding a random elastic distortion using a grid size of 10-by-10 pixels, shearing along the horizontal or vertical axis, and adding custom Poisson noise. Each operation has a 50% probability of occurring for each image input through the pipeline. For a six-fold augmentation using the pipeline, this resulted in a total 56-fold increase in training data. When splitting the data, 75% for training and 25% validation, care was taken to ensure an equal class balance between both sets of data.

### Loss function evaluation

Three prediction data sets were generated to simulate loss function response to true-positive, true-negative, false-positive and false-negative occurrences in a faux training case (Fig. 4.3). Prediction set 1 displayed translational overlap between the predicted and ground-truth case, with equal false-positive and false-negative occurrence. Here, BCE appeared to consistently generate higher loss values, except when the IoU score was at its lowest, where BCE gave a loss value of 2.66 compared to the distance map loss, which gave a value of 3.00. Dice and FTL ( $\alpha=0.4$ ) loss behave very similarly which is not surprising considering the  $\alpha$  parameter is close to 0.5 which would be an equal weighting between false-positive and false-negatives, which was equal in this simulated prediction set. Combo loss appeared to increase as IoU score increased, which made the loss function poorly suited for the segmentation network

training. Varying the  $\alpha$  parameter for FTL did not affect the loss value reinforcing the earlier statement that weighting the loss for either false-positives or false-negatives was meaningless for this simulated prediction data set given their equal occurrence.



**Figure 4.3.** Evaluating loss functions using faux simulated data sets with eight predictions of increasing intersection-over-union (IoU) score. **A.** Prediction set 1 uses simulated predictions of a square object translating until complete incidence over a ground-truth square. This set features true-positives, true-negatives, false-positives and false-negatives, with false-positive and false-negative occurrences equal throughout the eight faux predictions. BCE loss performs best with Dice loss and FTL ( $\alpha=0.4$ ) performing equally. Combo loss increases value with improving IoU, making it ill-suited for U-Net training. **B.** Prediction set 2 features an initial

*prediction scaled smaller than the ground-truth but colocalised over the central position. Predictions incrementally scale to match the ground-truth. There are no false-positive occurrences. BCE performs best with consistently highest loss value generation. Varying FTL  $\alpha$  between 0.1 and 0.9 for weighting towards false-positives and false-negative penalisation, respectively, shows the highest loss when focusing on false-negative occurrence. C. Prediction set 3 features an initial prediction larger than the ground-truth but colocalised with the central position. Incrementally, the faux prediction scales smaller to match the ground-truth. There are no false-negative occurrences in this prediction set. BCE performs best once again. Varying FTL  $\alpha$  between 0.1 and 0.9 shows heavier penalisation with higher loss values when  $\alpha$  values are low; coinciding with a weighted focus towards false-positives penalisation. BCE, binary cross entropy loss; DM, distance map loss; BF, binary focal loss; FT, focal Tversky loss.*

Prediction set 2 had scaling in the faux prediction, from small to matching with the ground-truth. As the prediction was never bigger than the ground-truth, there were no false-positive occurrences. BCE appeared to generate higher loss values as seen in prediction set 1, with FTL ( $\alpha=0.4$ ) and Dice performing comparatively. Combo loss increased as IoU increased as in prediction set 1. In this scenario, varying FTL  $\alpha$  between 0.1 and 0.9 showed that in the absence of false-positives, higher  $\alpha$  values which weight towards false-negative penalisation gave higher loss values, whilst lower  $\alpha$  values which weight FTL towards penalising false-positives saw a reduction in loss values.

Prediction set 3 had scaling in the faux prediction from small to matching with the ground-truth. As the prediction is never smaller than the ground-truth, there were no false-negative occurrences. BCE again performed best with the highest loss value generation. Combo loss still performed the worst. Varying the FTL  $\alpha$  from low to high saw a general reduction in the overall loss as there were zero false-negative occurrences.

Both predictions set 2 and 3 show that FTL loss can be used to target training loss to penalise false-negatives or false-positives. If similar weighting were required, FTL  $\alpha=0.5$  would perform equally to Dice loss.



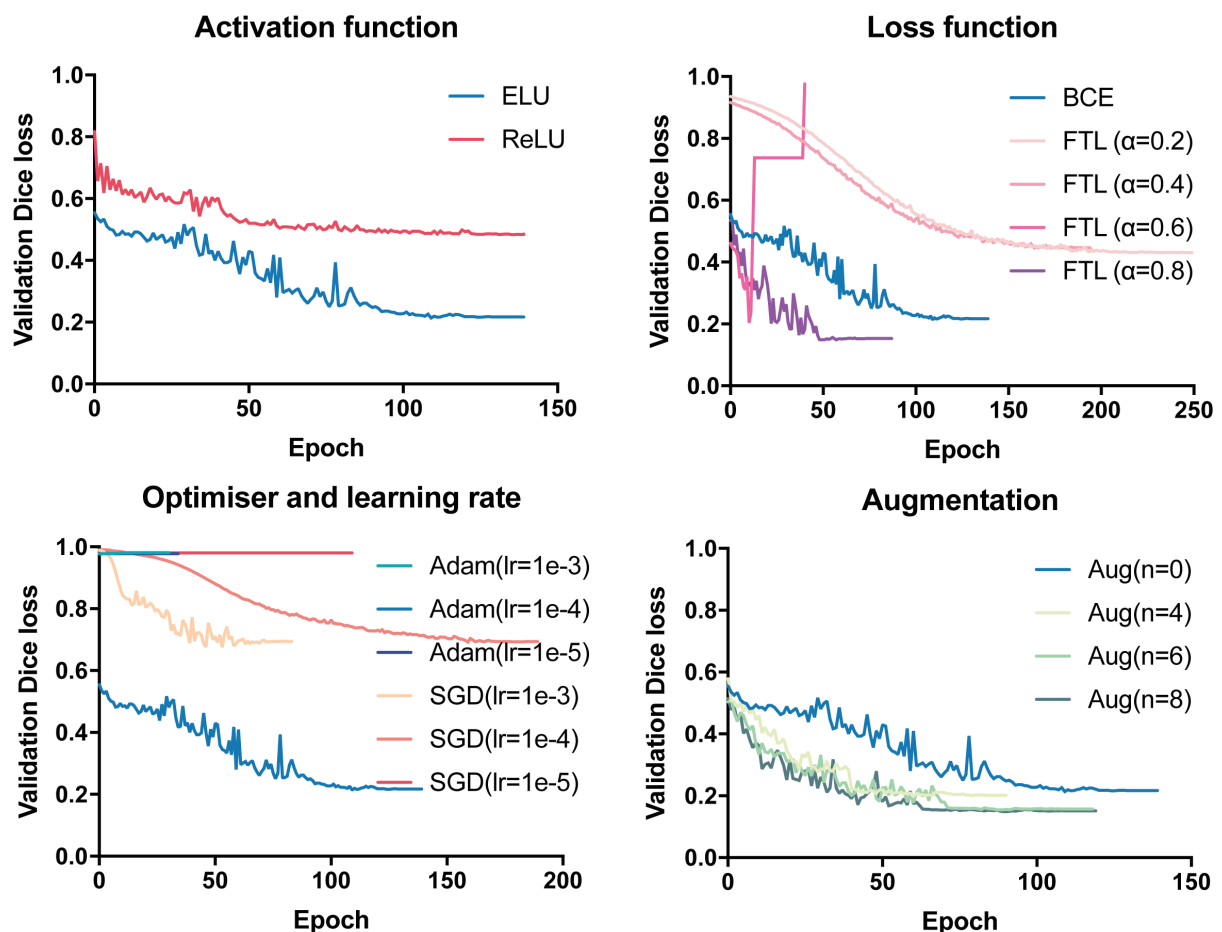
Of the evaluated loss functions, BCE and FTL provided the most advantageous loss functions to use in a real training scenario. BCE consistently generated higher loss values which would encourage network learning in all regimes where the network was predicting poorly (low IoU score) and in the regime where the network was performing well (high IoU score). This ensures a network would be consistently encouraged to optimise the weights without the loss being too large such as to shunt the network into drastically modifying the weights. FTL provided an interesting approach to target network learning towards minimising false-positives or false-negatives. As there was significant class imbalance (Fig. 4.2B) in the training data set, this could potentially force the network to learn more cell class features (or alternatively background class features) by modifying the  $\alpha$  parameter.

### **Hyperparameter tuning**

To get the optimum parameters for training the U-Net model, we ran several network training tests with variable hyperparameters. All scenarios were performed with the same conditions as specified in 'Methods' and evaluated using the Dice loss score. First, the activation functions ELU and ReLU were compared (Fig. 4.4). There was a notable improvement to the final validation loss reported during training when using the ELU activation function. Furthermore, ELU has been reported to improve the final network generalisation and also improve network training times (Clevert et al., 2015). ELU was therefore considered the activation function of choice for final U-Net training.

Following the results on evaluating various loss functions (Fig. 4.3), BCE and FTL functions were tested with the FTL  $\alpha$  parameter modified between 0.2 and 0.8 with increments of 0.2. BCE and FTL with  $\alpha=0.8$  provided the best validation loss. FTL with  $\alpha=0.6$  caused the network to diverge in performance whilst  $\alpha=0.2$  and  $\alpha=0.4$  converged to the same performance which was worse than the best BCE loss by a Dice loss measurement of 0.2, and worse than FTL with  $\alpha=0.8$  by a loss of 0.3. Whilst FTL with  $\alpha=0.8$  performed best, it also appeared to overfit after 50 epochs suggesting the model was overtraining on the training and validation data. This would suggest poor generalisation performance for the model. Consequently, BCE provided more robust training and was chosen for final U-Net training.

Varying optimiser and learning rate resulted in very stark differences between network training performance (Fig. 4.4). We observed the best performance with an Adam optimiser initialised with a learning rate of  $1e-4$ . Almost all other variations of learning rate and use of the SGD optimiser resulted in zero convergence of the model during training. An Adam optimiser with learning rate  $1e-4$  was therefore considered the best optimiser to use.



**Figure 4.4.** Hyperparameter tuning. To get the optimum parameter configuration for U-Net training, a series of test experiments were conducted using the same base model configuration of ELU activation function, BCE loss function, Adam optimiser initialised with learning rate  $1e-4$  and zero stage two augmentation. Firstly, the activation functions ELU and ReLU were compared, showing improved network performance with ELU. Given the other benefits of ELU, such as improved network training time and generalisation (Clevert et al., 2015), ELU provides an ideal activation function for network training. Next, the BCE and FTL loss functions were tested with additional varying of the FTL  $\alpha$  parameter between 0.2 and 0.8. BCE and FTL with  $\alpha=0.8$  gave the optimum network performance with other values of  $\alpha$  resulting in zero or poor

*convergence. FTL with  $\alpha=0.8$  appears to overfit after 50 epochs however suggesting it might not be ideal for training network to operate on novel data sets. When the optimiser and learning rate was modified, only the Adam optimiser with an initial  $1e-4$  learning rate showed network convergence. Finally, increasing the amount of stage two augmented data showed general improvement in network validation Dice loss compared to no stage two augmentation. After 6-fold augmentation, performance appears to saturate, suggesting zero gain for additional augmentation.*

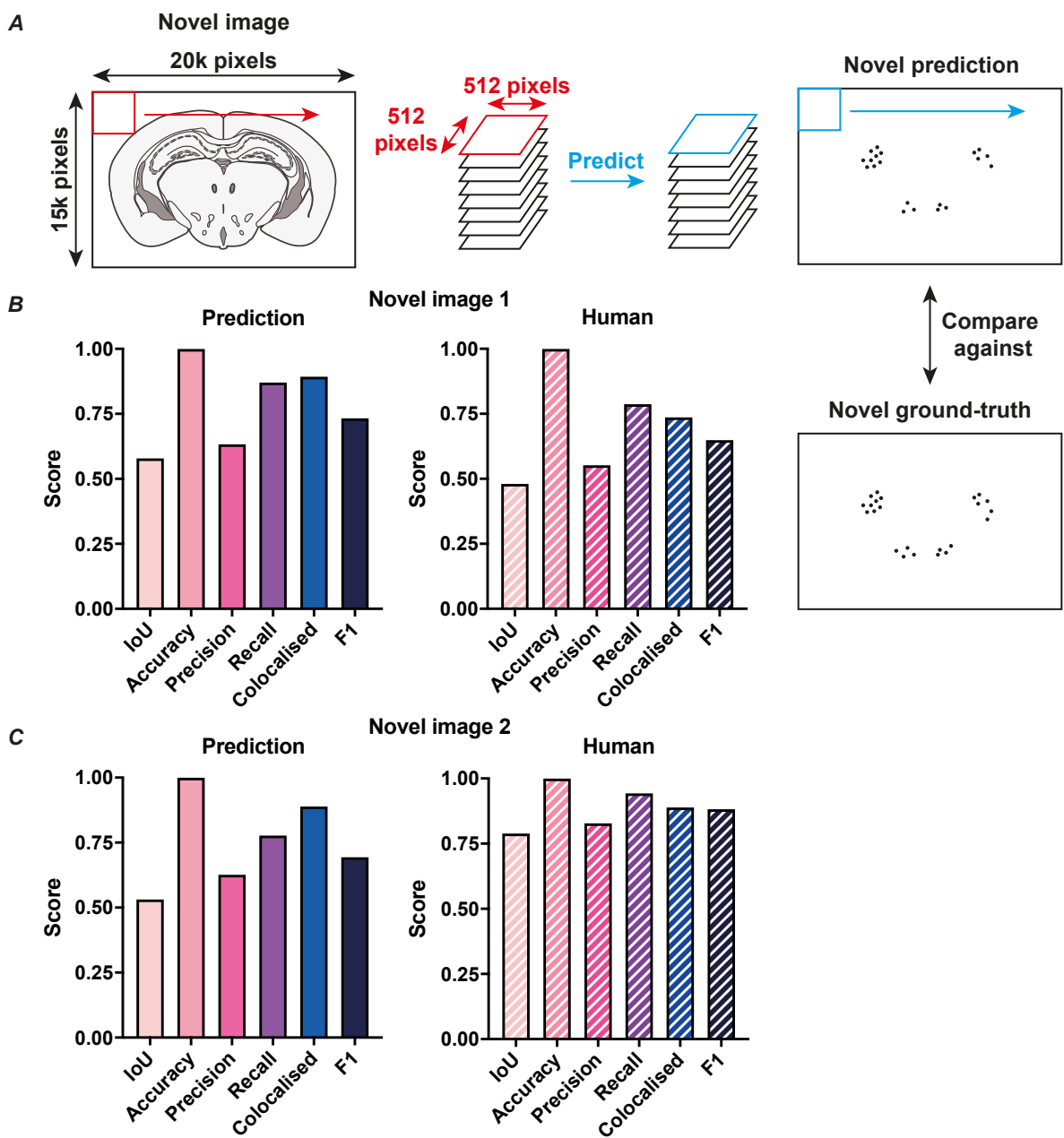
Lastly, the amount of stage two data augmentation, from 0-fold to 8-fold, was tested. Loss improved with additional augmentation data which was not surprising given that more training samples generally help improve network performance. There was no discernible improvement in Dice loss between six-fold or eight-fold augmentation at stage two, suggesting the gain in network performance saw minimal gain for high augmentation amounts. With this in mind, six-fold augmentation at stage two was chosen as the ideal number of training samples as it had fewer training samples and network training was, therefore, less memory consumptive and faster.

### **Optimised U-Net model training and evaluation**

Training of the final U-Net was done with the Adam optimiser initialised with a learning rate of  $1e-4$ , a BCE loss function, an ELU activation function, a batch-size of 8 and six-fold augmentation at stage two, giving a total training sample size of 12,264 which was split 75% (9,198 samples) for training data and 25% (3,066 samples) for validation data. Training completed after 18 hours and resulted in a final validation Dice loss score of 0.1542.

During inference on completely novel data (Fig. 4.5A), the trained U-Net predicted each image of  $20,000 \times 15,000$  pixels split into  $512 \times 512$  pixel blocks in 23 seconds using an NVIDIA Titan Xp GPU. For comparison, prediction using an Intel(R) Core(TM) i9-7980XE CPU took 4 minutes on each novel image, presenting a 10-fold improvement in speed with the GPU over the CPU. Prediction accuracy for the U-Net for novel image 1 (Fig. 4.5B) and novel image 2 (Fig. 4.5C) were both very high at 0.99982 and 0.99999, respectively. Both novel images resulted in IoU scores of 0.58 and 0.53 respectively, whilst precision and recall were higher with recall above 0.75 for both test images. However, these metrics do not often accurately reflect model

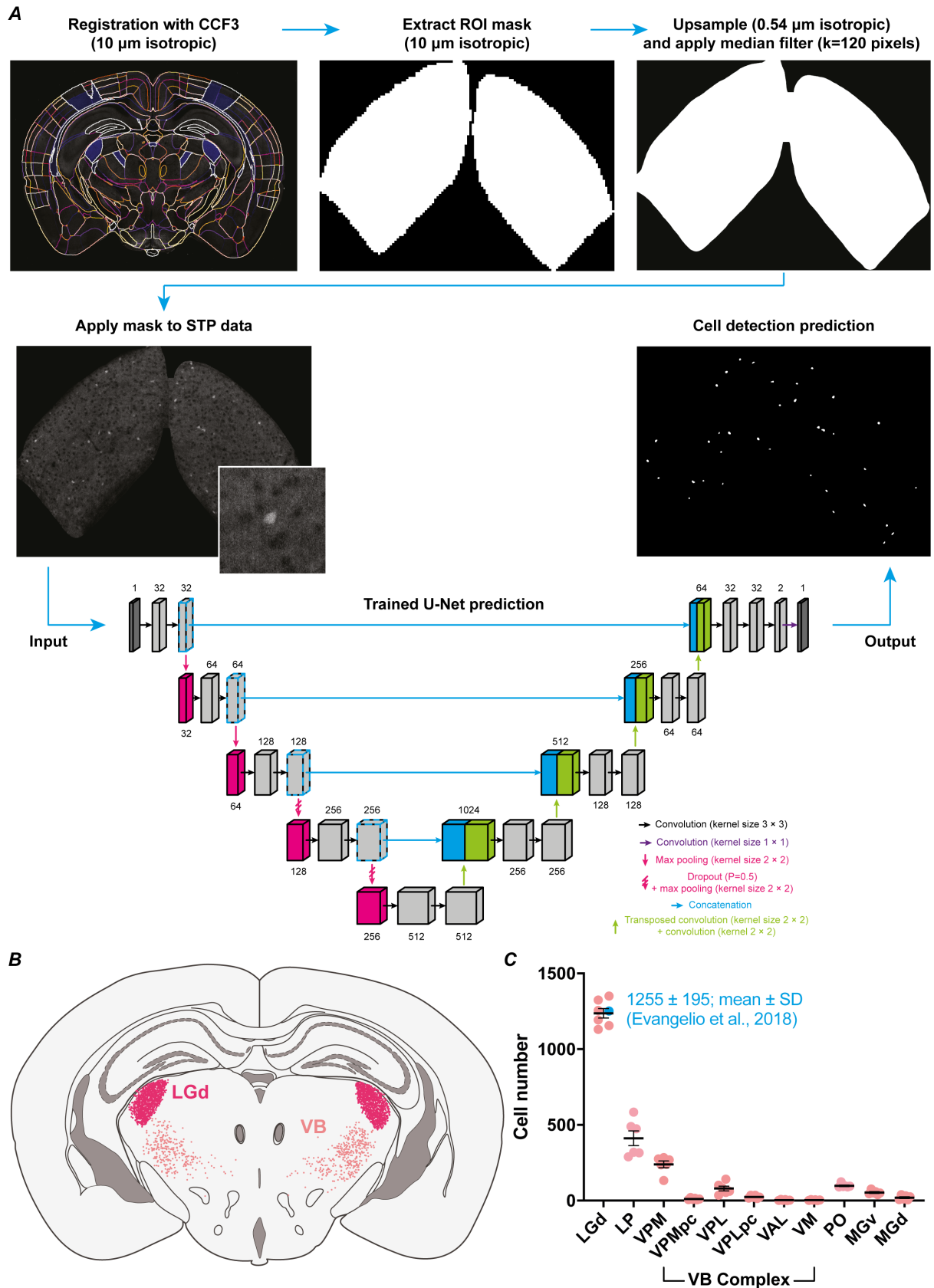
performance due to the high variability in segmentation result even if the challenge is easy. Instead, the colocalised metric shows a high hit-rate at 0.8929 and 0.8889 for novel image 1 and 2, respectively. When compared to a human rater, the U-Net outperformed in all evaluated metrics for novel image 1. For novel image 2, the human rater outperformed on all metrics except the colocalization metric, which had the same score at 0.8889. Compared to the U-Net prediction times, the human rater predicted novel image 1 in 2.5 hours and predicted novel image 2 in 30 minutes. Differences in human prediction times result from a different density of cell fluorescence in the novel images.



**Figure 4.5.** Trained U-Net evaluation using novel images. **A.** Novel images taken from data sets completely disconnected from data sets used for training the U-Net are split into  $512 \times 512$  pixel blocks. The image stack is predicted with the U-Net and reconstructed into the original image dimension for a one-to-one mapping between the novel STP images and the prediction maps. A threshold at 0.5 was taken to binarize the prediction map, with predictions  $>0.5$  classified as a cell and everything else classified as background. **B.** Prediction on novel image 1 is compared against the ground-truth and a separate human prediction. The U-Net performs better at every metric, including colocalisation, which measures the colocalised hit-rate of predicted objects to the ground-truth. **C.** Prediction on novel image 2 is compared against the ground-truth and a separate human prediction. The human prediction was almost universally better than the U-Net except for the colocalisation metric, which was equivalent at 0.8889. Arguably, the colocalisation presents the best metric to assess the U-Net model given that it does not penalise predictions for having slightly variable segmentation results but critiques the model on accurate detection of objects.

### **Anatomically targeted cell counting using trained U-Net**

For validation of the complete automated cell counting pipeline, we imaged the Sox14gfp/+ mouse (n=3) across the thalamus ( $0.54 \times 0.54 \times 10 \mu\text{m}$  XYZ-resolution) with automated cell counting targeted to the dorsal lateral geniculate nucleus (LGd), lateral posterior nucleus (LP), the ventrobasal complex (VB) and its sub-structures (ventral posteromedial nucleus, VPM; parvocellular part of the ventral posteromedial nucleus, VPMpc; ventral posterolateral nucleus, VPL; parvocellular part of the ventral posterolateral nucleus, VPLpc; ventral anterior-lateral complex, VAL; and ventral medial nucleus, VM), the ventral medial geniculate nucleus (MGv), the dorsal medial geniculate nucleus (MGd) and the posterior complex of the thalamus (PO) (Fig. 4.6).



**Figure 4.6.** Complete counting pipeline using registration to Allen CCF3 atlases to target automated cell counting to anatomical regions of interest. **A.** Results from registration to CCF3 atlas (10  $\mu\text{m}$  isotropic voxel size) allow masks for anatomical regions of interest to be

*extracted from STP data. Masks are upsampled to match original  $0.54 \times 0.54 \mu\text{m}$  voxel size and smoothed with a large median filter (kernel size 120 pixels) before applying to the STP data to extract fluorescence information. Extracted regions are passed through trained U-Net model for prediction of cells, which was thresholded at 0.5 to binarise the prediction. **B.** Example of automated counting in the LGd and VB complex in the thalamus of the Sox14GFP/+ mouse. **C.** Validation of automated cell counting on the Sox14GFP/+ mouse ( $n=3$ ) in the LGd at  $1,234 \pm 90$  (mean  $\pm$  SD) compared to a stereological study by Evangelio et al. (2018) which reported  $1,255 \pm 195$ . Other counts are shown for LP, VB complex (including sub-compartments PO, MGv and MGd), Dorsal lateral geniculate nucleus, LGd; lateral posterior nucleus, LP; ventrobasal complex, VB; ventral posteromedial nucleus, VPM; parvicellular part of the ventral posteromedial nucleus, VPMpc; ventral posterolateral nucleus, VPL; parvicellular part of the ventral posterolateral nucleus, VPLpc; ventral anterior-lateral complex, VAL; ventral medial nucleus, VM; ventral medial geniculate nucleus, MGv; dorsal medial geniculate nucleus, MGd; posterior complex of the thalamus, PO.*

The pipeline for one structure such as the LGd took 20 minutes to complete in one data set. Other structures would take longer or shorter, depending on their rostral-caudal and medial-lateral extent. In addition, multiple images were processed in parallel using a synchronised queue system processed by a Flask hosted U-Net model. Whilst parallelisation was possible, an improvement in processing time depends on how extensive the structure list for processing was, plus the resources available on the workstation. Images were only processed from the queue if a predefined amount of system resources were available to process it. Furthermore, as the STP data sets were terabyte in size and image processing is a computationally expensive, care was taken to choose the optimum Python packages to reduce image loading speeds and processing with Gaussian and median filters (Supplementary Fig. 4.1).

The overall pipeline enabled automated counting in target structures (Fig. 4.6B) with minimal human interaction. The count of Sox14 expressing cells in the LGd of P21 Sox14gfp/+ mice returned a value of  $1,234 \pm 90$  (mean  $\pm$  SD,  $n=3$  mice) Sox14 expressing cells (Fig. 4.6C). Here Sox14 cells are considered a marker for local interneurons in the LGd as previously reported (Jager et al., 2016). This count was validated against a manual stereological study by Evangelio

et al. (2018) which reported  $1,255 \pm 195$  (mean  $\pm$  SD) GABAergic interneurons in the wild type adult C57BL/6 mouse (Fig. 4.6C), demonstrating that our pipeline was agreeable in the count with a much-reduced variance. Automated counts (mean  $\pm$  SD) in other thalamic areas were  $411 \pm 109$  in the LP,  $251 \pm 49$  in the combined VPM and VPMpc,  $105 \pm 40$  in the combined VPL and VPLpc,  $4 \pm 3$  in the VAL,  $4 \pm 2$  in the VM,  $99 \pm 12$  in the PO,  $53 \pm 12$  in the MGv and  $20 \pm 13$  in the MGd. All reported numbers are for a single hemisphere and were averaged across both hemispheres (n=3 mice).



#### 4.4 DISCUSSION

Here, we have presented a novel, fast, automated cell detection approach with demonstrated use on whole-brain STP data sets using registration with the Allen CCF3 atlases to target quantification to anatomical regions of interest. Robust model evaluation through novel image prediction, simulating a deployment case, proves generalisation of the network. Pipeline quantification of interneurons in the LGd of the Sox14GFP/+ mouse ( $1,234 \pm 90$  cells,  $n=3$  mice) shows remarkable agreement with an independent manual stereological study by Evangelio et al. (2018) ( $1,255 \pm 195$  cells). Whilst the LGd was quantified in 20 minutes for a single sample, the pipeline was constructed in a way to process multiple structures in parallel. Therefore, execution time was dependant on the number of structures and computational resources of the system. Overall, the most significant bottleneck in the pipeline was the prediction using the U-Net. TensorFlow and Keras are not optimised for multi-threading due to how they configure a systemwide environment on instantiation. We expedited the pipeline with parallel processing, by using a synchronised queueing system in combination with hosting the U-Net model on a multi-threaded Flask server. Individual inference environments with Flask prevented conflicts between multiple instances of TensorFlow during parallel U-Net predictions. Due to the size of the data, the pipeline was still restricted based on the number of resources available. Flask hosting of the network works but does not necessarily present the optimum approach to parallelise the system.

The fastest inference was observed when using the GPU at 23 seconds for a  $20,000 \times 15,000$  pixel image, split into  $512 \times 512$  pixel blocks, and predicted in batch. The batch prediction used here was native to Tensorflow and Keras but was not possible to be used for batch prediction for parallel processing of structures and images. This is because STP images are processed one-by-one to significantly reduce RAM requirements. Images passed to the U-Net for inference were also different dimensions which makes them incompatible for batch processing without making all images in the batch of equal size, which could limit RAM requirement. However, whilst the GPU was able to utilise multiple processing cores for fast inference, RAM requirements were limited meaning images much larger than  $512 \times 512$  pixels were unable to be predicted by the GPU and had to be done using the CPU.

During pipeline optimisation, it was found that many commonly used Python image packages did not necessarily lead to the fastest result (Supplementary Fig. 4.1). For example, Pillow (PIL) package offers one of the most ubiquitously used image libraries, and the SIMD fork (Pillow-SIMD) offers an even more optimised performance. However, image reading time was higher at 6 seconds for a single  $20,000 \times 15,000$  STP image, compared to Tifffile which took 2 seconds when tested on a macOS 10.14.5 Intel(R) Core(TM) i5 at 3.30GHz workstation. The speed difference for the same package versions was negligible on an Ubuntu Intel(R) Core(TM) i9-7980XE CPU at 2.60GHz workstation. However, the 4 seconds difference between the two image libraries would result in a total difference of 66 minutes when loading 1000 images. As entire STP data sets contain a thousand images, a saving of one hour in processing time is not negligible, especially considering this is only loading the data prior to processing.

Image filtering is a common component of image processing. Gaussian filtering is most effective for removing image noise whilst median filtering is particularly effective at maintaining boundaries between structures in an image. In the pipeline, median filtering was used to smooth the upsampled mask boundary into a more anatomically accurate delineation. With a very high upsampling factor from  $10 \mu\text{m}$  isotropic to  $0.54 \mu\text{m}$  isotropic, the median kernel was required to be very large at a size of  $120 \times 120$  voxels to smooth a large dimension image appropriately. Unfortunately, median filtering is computationally slow as it requires a sorting algorithm to determine the median value. OpenCV had the fastest median filtering implementation taking 2 seconds to apply a  $120 \times 120$  kernel to a test image, compared to the Pillow-SIMD and Scipy libraries which took 30 seconds. Putting this into perspective, if only analysing a single structure over the 1,000 images, using OpenCV for image filtering operations would save 470 hours (20 days) of processing time. Due to the sheer volume of data being processed here, these optimisation steps were crucial for maintaining a reasonable processing time. Powerful workstations are not always easily accessible. The automated pipeline presented here was therefore designed in a way which would be as minimally demanding on computational resources. It performed equally well on a macOS 10.14.5 Intel(R) Core(TM) i5 at 3.30GHz workstation and a more powerful Ubuntu Intel(R) Core(TM) i9-7980XE CPU at 2.60GHz workstation. A GPU with more memory would,

however, massively expedite processing time as the most significant bottleneck was image inference with the network.

## 4.5 FUTURE PERSPECTIVES

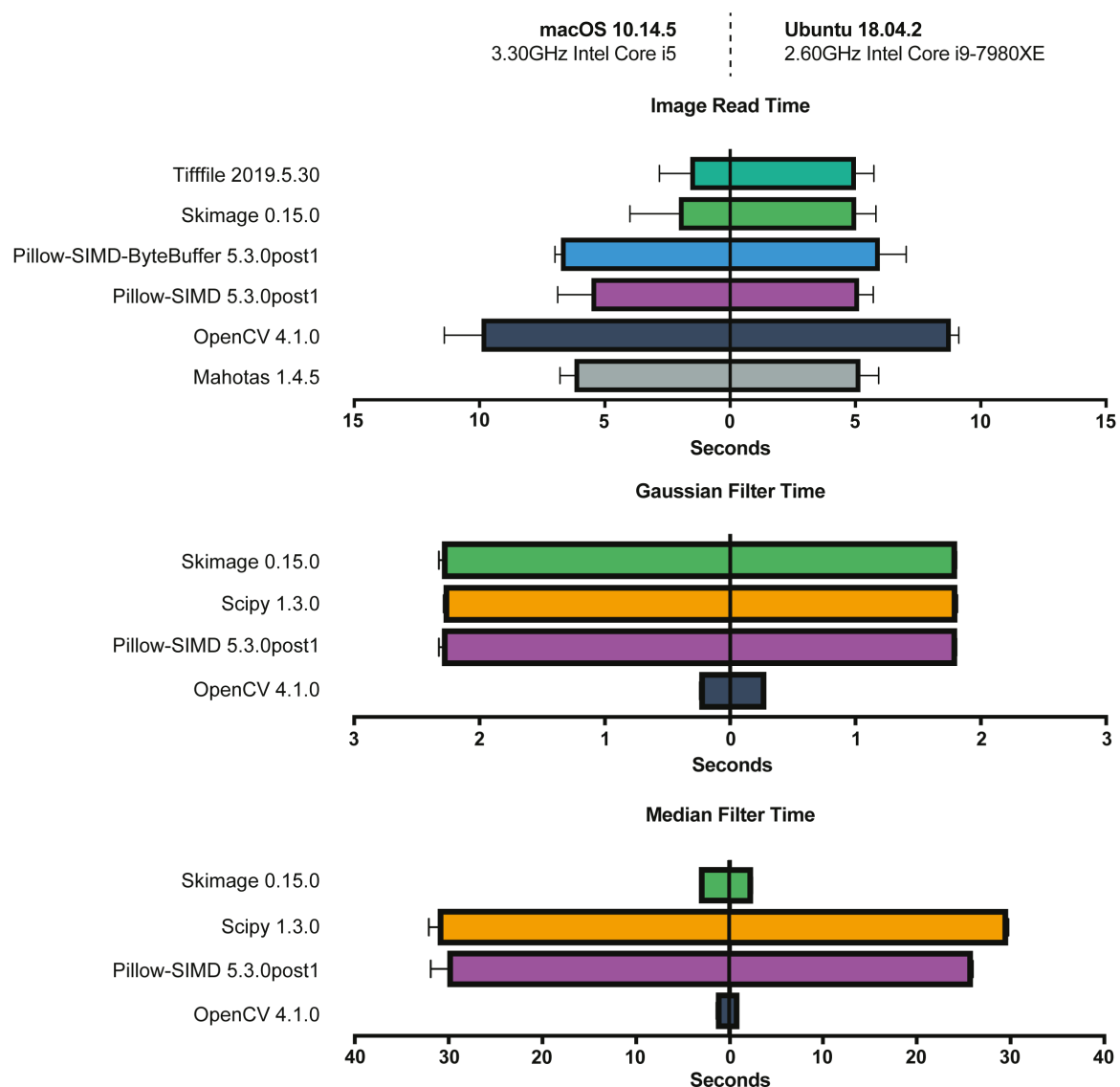
Performance of a U-Net model, and for general image classification network, comes down to the amount of training data. As we have shown here, image augmentation can be effective at bolstering the number of training samples and can even be used to improve network generalisability by augmenting with additive noise and elastic distortions. However, augmented data is only a partial surrogate for real data. Simple transformations are useful for making networks rotation- or scale-invariant. However, learning of additional features can only be done if the data set is increased with new data which contains new features. For example, training data presented here comes from soma localised fluorescence. The U-Net would, therefore, be poorly optimised if presented with nuclei localised fluorescence for inference. Problematically, manual segmentation of data is not easy and currently no comprehensive database exists of manually segmented cells from a range of different imaging experiments. This would be hugely beneficial for training networks to be more generalised.

That said, additional methods of augmentation could play a pivotal role in improving network performance. ConfocalGN (Dmitrieff & Nédélec, 2017) presents a method to generate synthetic microscopy image stacks using a ground-truth image stack. It works by using a reference stack to calculate noise properties, before convolving the ground-truth with a point-spread-function, extracting a new image and then adding noise back into the data. Whilst the technique was only explored on a confocal stack of a microtubule and is partially limited in its ability to generate background information, such a process could prove valuable for gathering additional training data. Recent efforts have also gone towards another class of machine learning called generative adversarial network (GAN; Goodfellow et al., 2014). These networks compete two smaller neural networks against one another. One network generates images using learned features from the training data set, whilst the other network acts as a discriminative classifier which attempts to classify the generative output and encourages the generative neural network to improve. Therefore, it would be interesting to explore the use of GANs to generate synthetic microscopy data in addition to a corresponding segmentation mask.

One way to improve the pipeline execution time would be to downsize the data passed into the network for inference. The original STP acquisition spatial resolution of  $0.54 \times 0.54 \mu\text{m}$  is

ideal for mapping fine structures such as dendrites and axons. However, for quantifying soma signal, it might be possible to downsize the data to a spatial  $1 \times 1 \mu\text{m}$  resolution which would still retain signal from the soma whilst being a quarter of the original data size. Multiple successive coronal image planes could also be maximum intensity projected or averaged over the Z-axis as a means for pseudo-averaging. This would improve the signal-to-noise level and possibly improve detection accuracy. Depending on the way this is implemented, it might also reduce the amount of data to predict, thereby improving processing time.

## 4.6 SUPPLEMENTARY DATA



**Supplementary Figure 4.1.** Performance times for a variety of Python image processing libraries for image reading, Gaussian filter operation and median filter operation. Performance was tested on a macOS 10.14.5 Intel(R) Core(TM) i5 at 3.30GHz workstation (left) and an Ubuntu Intel(R) Core(TM) i9-7980XE CPU at 2.60GHz workstation (right). The image read times for a  $20,000 \times 15,000$  pixel STP image were comparable between the two workstations except for TiffFile and Skimage libraries which were around 3 seconds faster to load an image in the macOS workstation. TiffFile offers the fastest image reading time. OpenCV performed the best for filter operating including the Gaussian and median filter when executed with a kernel size of 120 pixels. OpenCV performed drastically faster than the other alternatives tested. Error bars represent standard deviation.

**INTERNEURONS WIDELY POPULATE MANY NUCLEI IN THE MOUSE THALAMUS AND ARE CONNECTED TO THE RETICULAR NUCLEUS AND OTHER INTERNEURONS OF THE VISUAL THALAMUS**

**ABSTRACT**

Local interneurons establish 10% of the neuronal population in the dorsal lateral geniculate nucleus (LGd) and are required for strong inhibitory modulation of visual information transmission to the primary visual cortex (VISp). The other resident cells, thalamocortical (TC) relay cells, receive input from the retinal ganglion cells (RGCs), the reticular nucleus (RT), VISp layer 6, superior colliculus (SC), and from the local interneurons. However, the mapping of inputs onto LGd interneurons remains relatively unexplored. To this end, we performed high-resolution serial two-photon tomography (STP) on the Sox14 transgenic mouse to identify the thalamic interneuron population. We counted a total of  $6,588 \pm 811$  Sox14 cells ( $n=3$  mice) across the complete thalamus with higher numbers observed in sensory and motor-related areas and further prevalence in first-order nuclei. This result demonstrates that a more significant number of interneurons exist outside of the visual thalamus areas than previously expected (Arcelli et al., 1997). We used a modified G-deleted rabies virus to map monosynaptic inputs onto interneurons of the LGd. Strong input from the RT and SC established previously unseen connectivity and evidence of tectogeniculate input. Patch-clamp recording of LGd interneurons following optogenetic stimulation of RT GABAergic neurons confirmed the existence of functional connectivity from the RT. Simultaneous patch-clamp recording from interneuron pairs also revealed functional interneuron-interneuron connectivity. Imaging of biocytin filled interneuron pairs revealed extensive morphological complexity with dendritic trees spanning a volume of almost  $600 \times 600 \times 250 \mu\text{m}$ . Close ( $<5 \mu\text{m}$ ) appositions between putative dendrites and axons indicated preferential connectivity within a  $100 \mu\text{m}$  radius of the soma (39 of 72 measures; 54%) over connections in the distal regions of the dendritic tree  $>100 \mu\text{m}$  from the soma (33 of 72 measures; 46%). High tortuosity along thicker dendrites and the spindled morphology of recently termed targeting dendrites suggested multiple terminal contacts at what a likely to be specialised glomerular synapses. The evidence presented in this study shows that interneurons have a high

morphological complexity with far-reaching dendrites which make them well suited for local and distal inhibitory modulation of TC relay cells guided by inputs such as the RGC, corticogeniculate projections, tectogeniculate projections, other local interneurons and surprisingly, from GABAergic RT neurons.



## 5.1 INTRODUCTION

The thalamus is a diencephalic brain region situated dorsal to the hypothalamus and is cranially encapsulated by the telencephalon and is posteriorly adjacent to the mesencephalon. Classically, the structure is seen as a relay centre for almost all corticopetal (projecting to the cerebral cortex) sensory information - the exception is olfaction which relays information to the primary olfactory cortex through the olfactory bulb (Courtiol & Wilson, 2015). The centralised position of the thalamus makes it strategically positioned to modulate all sensory information as it passes to the cortex and the presence of dense cortico-thalamic and reticular-thalamic feedback provides further complexity to sensory processing in the thalamus.

Neurons cluster into specialised sensory nuclei which are distributed across the dorsal, intermediate and ventral tiers of the thalamus (Schröder et al., 2020). The dorsal tier encloses the majority of thalamic nuclei and is mostly comprised of thalamocortical (TC) relay cells. The intermediate tier appears to primarily be involved in somatosensory information although knowledge of this region of the thalamus is relatively sparse. The ventral tier consists of the medial geniculate nucleus (MG) which serves as an auditory relay (Hackett et al., 2011) as well as the ventral lateral geniculate nucleus (LGv), intergeniculate leaflet (IGL) and also the thalamic reticular nucleus (RT; Schröder et al., 2020). The RT is most heavily involved in modulating TC relay function of primary sensory nuclei in the dorsal thalamus through a set of GABAergic neurons that sit in a shell encapsulating the rostral and lateral thalamus.

In addition to RT inhibition, local GABAergic interneurons are distributed across relay nuclei in the dorsal thalamus. However, their abundance and thalamic distribution is species dependant. In rodents, local interneurons are confined to the LG with a local interneuron population of around 15-20% (Arcelli et al., 1997; X. Wang et al., 2011). The cat also has an LG interneuron population of 20-25% (LeVay & Ferster, 1979). In the monkey, this number is higher at 25-35% (Montero, 1986). For primates, the thalamic interneuron population outside the LG is still high at 20-30%, but interneurons numbers are considered to be much lower in the rodent species (Arcelli et al., 1997). Reduced interneuron number in the thalamus has led to speculation that their population could be representative of species complexity.

Of the numerous sensory roles of the thalamus, the thalamic visual processing network is perhaps the most extensively studied at both the anatomical and functional levels. Retinal ganglion cells (RGCs) project to retinorecipient nuclei in the thalamus, including the dorsal lateral geniculate nucleus (LGd), intergeniculate leaflet (IGL), ventral lateral geniculate nucleus (LGv), lateral posterior nucleus (LP), centrolateral nucleus (CL), para-habenular zone (PHb), peripeduncular nucleus (PP), anterodorsal nucleus (AD), zona incerta (ZI), subgeniculate nucleus (SubG) and superior colliculus (SC) (Lawrence & Studholme, 2014). Of these retinorecipient nuclei, the LGd, IGL and LGv receive abundant RGC input with the LGd estimated to receive input from 37% of RGCs, but strikingly only amount to 10% of the synapses (Bickford et al., 2010; Martin, 1986). The LGv is split into two laminae; the external LGv which receives dense RGC input, and the internal zone which receives none (Niimi et al., 1963). Despite RGC axons traversing through the LGv to target the LGd, many RGC classes which innervate the LGd do not innervate the LGv, meaning both nuclei receive segregated RGC class input. The LGd, IGL and LGv are functionally distinct. The LGd primarily relays image-forming information to the primary visual cortex (VISp), whilst the LGv and IGL are required for non-image-forming and associated visual information processing, including visuomotor coordination (Ciftcioglu et al., 2020) and circadian entrainment (M. E. Harrington, 1997). A large number of LGv and IGL projections, therefore, additionally target the superior colliculus (SC) which integrates multisensory information such as visuomotor function. Relay cells of the LGd are clustered into three classes; X, Y and W. Each class has a distinct morphology and localisation within the LGd with X-cells displaying a bipolar dendritic tree and are located in the ventral LGd, Y-cells displaying a spherically spanning tree and occupy the central core of the LGd, and lastly, the W-cells showing a hemispherically spanning dendritic tree and occupy the LGd shell (Guido, 2018).

In terms of LGd afferent connectivity, glutamatergic TC relay cells receive 5-10% of their inputs from glutamatergic RGCs of the retina, yet RGC input is considered the primary excitatory drive of the LGd despite this low percentage (Bickford et al., 2010; Sherman & Guillery, 2002). RGC input is spatially arranged in the LGd with ipsilateral RGC input clustered into the ventromedial core of the LGd with contralateral input surrounding the ipsilateral core in a non-overlapping shell (Kerschensteiner & Guido, 2017). The LGd of rodents preferentially receives contralateral eye input with over 95% of RGC axons originating from the contralateral

eye, compared to 60% in humans (Petros et al., 2008). RGC inputs are also topographically arranged, meaning RGCs carrying spatially localised information from the visual field innervate spatially localised TC relay cells. Spatial information in the visual field is, therefore, faithfully reconstructed in the thalamus. Of the 35 or so RGC classes that exist in the rodent (Baden et al., 2016), not all classes innervate the LGd. Some direction-selective RGC classes appear to preferentially target subdivisions of the LGd (Cruz-Martín et al., 2014). In addition, as many as 80% of the RGC classes which target the LGd also appear to target the SC (Ellis et al., 2016). Some TC relay cells might also receive RGC input from a vast number of RGC classes as it appears classes of RGC axon terminals spatially overlap across the LGd (Rompani et al., 2017).

The organised mapping of RGC inputs to LGd appears to be established during postnatal development. Axons from the contralateral eye form inputs to the LGd in the later stages of embryonic development (E16-P0). Conversely, axons from the ipsilateral eye establish their inputs postnatally (P0-P2). Arriving first, the contralateral RGCs arbour spread throughout the LGd with the later arriving ipsilateral arbour creating an overlapping region in the medial dorsal LGd. Over the first two postnatal weeks, RGC arbour undergo pruning causing the overlapping ipsilateral and contralateral domains to segregate into the developed non-overlapping organisation seen in adult mice (Jaubert-Miazza et al., 2005). During this development period, the TC relay cells of the LGd appear to already function as mature cells.

The rest of the synaptic input to TC relay cells come from non-retinal sources with around 50% of those originating from VISp layer 6 (Sherman, 2004), and other inputs from the thalamic reticular nucleus (RT), local interneurons, superior colliculus (SC) and brain stem. VISp corticothalamic projections form their synaptic contacts on the more distal dendrites of the TC cell but do not appear to show distinct domain preferences. Therefore, it is thought these cortical feedback loops are primarily for modulatory purposes through weak inputs to the TC relay cells, and do not serve to influence the receptive field properties of the LGd. It is also believed that the weak corticothalamic inputs could be required for controlling TC firing pattern between burst or tonic behaviour (Sherman, 2016). Deconstructing their modulatory function is, however, challenging due to the complex, class-specific arrangement of corticothalamic inputs into the LGd (Tsumoto & Suda, 1980). In contrast to the modulatory

role of VISp inputs, tectogeniculate projections from the SC appear to synapse on the proximal dendrites of TC relay cells and provide driver-like inputs. As the driver to TC relay cells come from RGCs, this suggests a notion that thalamic neurons might possess multiple drivers, or driver-like inputs (Bickford et al., 2015).

The other significant input to the TC relay cells of the LGd come from GABAergic reticular neurons of the RT which are recruited by axon collaterals of TC relay cells coming from the ascending thalamocortical pathway to VISp layers 1 and 2/3 for shell located W-cell axons and layer 4 for the more centrally located X and Y-cells. Additionally, descending VISp layer 6 corticothalamic projection also send collaterals to the RT on the way to the LGd (Ohara et al., 1980). Indeed, the interplay between excitatory and inhibitory inputs to TC relay cells are highly complex.

The resident GABAergic interneurons are often overlooked but are functionally and morphologically compelling. Amounting to 20-30% of the neuronal population of the rodent LGd, the local interneuron role is to provide strong feedforward inhibitory input of TC relay cells to modulate thalamocortical driver information transmission (Guido, 2018). Fate mapping studies have shown that interneurons of the LGd originate from two progenitors; the ventral wall of the third ventricle (Golding et al., 2014) and the tectum (i.e. the SC)(Jager et al., 2016). During the first two weeks of postnatal development, interneurons migrate into the LGd and are guided by retinal inputs to establish their distribution and functional integration with TC relay cells (Golding et al., 2014). Until maturity is reached, and much like the relay cells, interneurons show evidence of dendritic pruning but still display far-reaching dendritic morphology which remains within the nucleus (Morgan & Lichtman, 2020).

Local interneurons form two types of synaptic terminal with all classes of TC relay cells; axodendritic synapses (F1) form between the interneuron axon and the TC cell dendrites, and dendrodendritic synapses (F2) which form between the interneuron dendrites and TC cell dendrites through pre- or post-synaptic terminals (Sherman, 2005). Not all interneurons have axons meaning they only inhibit relay cells through F2 terminal sites on their dendrites. Accordingly, F2 terminals are the most frequently observed synapse between local interneurons and TC cells (Sherman & Guillery, 2006). F2 terminals can also form complex tri-

synaptic arrangements with RGC axon boutons and a relay cell. Although a synaptic bundle like this is more generally known as a glomerulus, the name triad most commonly refers to the tri-synaptic arrangement which encompasses a synapse between RGC axon and dendritic F2 terminal of the interneuron, with a second synapse between the RGC axon and a relay cell dendrite, and a third synapse between an F2 terminal of the interneuron and the relay cell dendrite (simplified as  $\text{RGC} \rightarrow \text{IN}[\text{interneuron}] \rightarrow \text{TC}$  with  $\text{RGC} \rightarrow \text{TC}$ ), all under the same glomerulus formation. However, increasing evidence shows that the triadic arrangement is not only reserved for the RGC driver input. Instead, the triadic motif has also been shown to occur with a relay cell axon in place of the RGC axon ( $\text{TC}_1 \rightarrow \text{IN} \rightarrow \text{TC}_2$  with  $\text{TC}_1 \rightarrow \text{TC}_2$ ) (Bickford et al., 2008), between RGC and interneurons only ( $\text{RGC} \rightarrow \text{IN}_1 \rightarrow \text{IN}_2$  with  $\text{RGC} \rightarrow \text{IN}_2$ ), between interneurons and TC cells only ( $\text{IN}_1 \rightarrow \text{IN}_2 \rightarrow \text{TC}$  with  $\text{IN}_1 \rightarrow \text{TC}$ ) and with other unknown cells ( $\text{U}[\text{unknown}] \rightarrow \text{IN} \rightarrow \text{TC}$  with  $\text{U} \rightarrow \text{TC}$ ) (Morgan & Lichtman, 2020). Complex triadic formations are thought to be necessary for allowing dendritic branches of interneuron to be functionally independent for widespread and focal inhibition.

F1 terminals may also be located inside a glomerulus, although importantly they do not share in the triadic arrangements. In contrast, GABAergic projections from the RT do not appear to contact the dendrites of TC relay cells within glomeruli, but prefer to form their contacts on the distal dendrites (S. Wang et al., 2001). Independent F1 and F2 terminal function allows interneurons to intricately influence the processing of retinal information and signals that are transmitted to the cortex. The complex and far-reaching morphology of interneurons also allows them to innervate as many as 256 TC relay cells, spread across functionally segregated regions of the LGd, revealed following comprehensive reconstruction of a single interneuron (Morgan & Lichtman, 2020). The same study also showed that an interneuron dendritic tree could span a 500  $\mu\text{m}$  radial volume. Neurite structures could be split into trunk dendrites as classified by thick, long and directed neurite morphology, targeted dendrites as classified by thinner and more tortuous morphology which arborised from the trunk dendrites, and lastly the axon neurite which was classified based on lack of RGC terminal input but otherwise similarly shaped as the more directed trunk dendrites. Three classes of neurites further exemplify that a single interneuron can have multiple functional properties to innervate surrounding TC cells in the LGd. Whilst it was not possible to confirm the origin of the unidentified cells establishing terminals with the single interneuron in the study, either via

monadic, dyadic or triadic motifs, these cells were thought to originate from the brainstem, pretectal region as evidenced in the cat (S. Wang et al., 2002), and the visual cortex as also evidenced in the cat (A. J. Weber et al., 1989). The work presented by Morgan & Lichtman (2020) is comprehensive and impressive as an endeavour. However, there is still more to glean from interneuron connectivity and their complex inhibitory role. The unidentified inputs could derive from tectal regions as already seen with SC driver-like input onto TC relay cell (Bickford et al., 2015) or as corticothalamic input from layer 6 VISp. There is also a question of interneuron-interneuron connectivity, a mechanism which has already been observed in primates (Pasik et al., 1976). There are several mechanisms involved during the development and maturation of the LGd, particularly concerning the local interneurons. Targeting dendrites which seek out specific synaptic arrangements (Morgan & Lichtman, 2020), in combination with retinal-input-driven interneuron distribution across the LGd (Golding et al., 2014), alludes to a high level of connectivity and complexity amongst interneurons.

To this end, we performed an investigation into interneuron connectivity by using the Sox14 gene as a marker for interneurons in the LGd (Jager et al., 2016). We mapped the Sox14 interneuron population across the entire thalamus using serial two-photon (STP) tomography in combination with cell count quantification using a custom deep-learning pipeline (see 'Chapter 4'). We incorporated anatomical delineation according to the Allen Brain Atlas (see 'Chapter 3') to target and quantify counts in all sub-structures of the thalamus. We also performed monosynaptic input mapping of interneurons using a G-deleted rabies virus tracing method to reveal monosynaptic inputs to the interneuron population. Lastly, we looked at interneuron-interneuron morphology and connectivity using a deep-learning neuronal tracing technique to aid paired-interneuron reconstruction alongside *in-vitro* patch-clamp functional recordings.

## **5.2 METHODS**

### **Animals**

The various Sox14 transgenic mouse lines were derived in King's College London as detailed in Jager et al. (2016). The Sox14gfp/+ transgenic line tagged Sox14 positive cells with GFP, the Sox14gfp/gfp transgenic line knocked-down Sox14 expression with double GFP transgene expression, and the Sox14Cre/+ transgenic line had Sox14 positive cells expressing the Cre recombinase transgene for targeting with the rabies virus tracing. All lines were maintained on a C57BL/6 background by Polona Jager, King's College London.

### **Monosynaptic presynaptic interneuron labelling**

For priming starter cells for rabies uptake, 50 nl of rAAV2-EF1a-DIO-TVA-G and 25 nl of rAAV2-EF1a-DIO-mCherry (2:1 ratio) was co-injected into the LGd of P18 Sox14Cre/+ mice. Coordinates for stereotaxic injection were determined using littermates, with an average injection coordinate at AP 2.26, ML 1.75 and DV 2.75. Mice were kept under anaesthesia with isoflurane for the duration of the procedure. After 2 weeks, SAD-rabies-dG-GFP (EnvA) was injected into the same coordinate location using the same injection procedure. After 9-10 days, mice were sacrificed as detailed below. The rabies injection procedure was done by Polona Jager, King's College London.

### **Sample preparation for STP imaging**

All mice for imaging with STP tomography were anaesthetised with IP injection of sodium pentobarbital (50 mg/kg) and sacrificed with transcardial perfusion with 20ml 1X phosphate-buffered saline (PBS) at 4°C, followed by 20ml 4% paraformaldehyde (PFA) at 4°C. Brain tissues were post-fixed in 4% PFA and kept at 4°C for 24 hours before transferring to 1X PBS with 0.01% NaN<sub>3</sub> for preservation. Prior to imaging, samples were embedded in a 4.5% oxidised-agarose solution as described in 'Chapter 2'.

### **Whole-brain STP tomography**

All mice were STP imaged using TissueCyte 1000 as described in 'Chapter 2'. For GFP imaging in the Sox14gfp/+ and Sox14gfp/gfp mice, 920 nm excitation wavelength was used in combination with GFP emission collected on channel 2 with a 525/40 nm Brightline bandpass filter (Laser 2000 Ltd.). We chose an excitation wavelength of 920 nm for imaging rabies

injected Sox14cre/+ mice. GFP emission was collected on the green channel with a 525/40 nm Brightline bandpass filter (Laser 2000 Ltd.), and mCherry emission was collected on the red channel with a 600/37 nm Brightline bandpass filter (Laser 2000 Ltd.). Image tile data was stitched following the procedure detailed in 'Chapter 2'.

### **Automated cell counting**

Cell counting was achieved using the method outlined in 'Chapter 3'. Following registration of STP data to the Allen Common Coordinate Framework version 3 atlases (CCF3), we targeted cell counting to delineated regions of interest using a deep-learning approach with custom trained U-Net network, with a previously validated performance (see 'Results' of 'Chapter 3'). For all thalamic cell counting in Sox14gfp/+ (P21, n=3) and Sox14gfp/gfp mice (P21, n=2), the pipeline was targeted to the Thalamus (TH), covering the sensory-motor related nuclei (DORsm) and polymodal association nuclei (DORpm), plus all sub-nuclei.

For automated counting in rabies traced Sox14cre/+ samples (3 months, n=3), the automated cell detection pipeline was targeted to the geniculate group of the dorsal thalamus (GENd), reticular nucleus (RT), anterior group of the thalamus (ATN), lateral group of the thalamus (LAT), the ventral group of the thalamus (VENT), and finally the primary visual cortex (VISp).

### **Electrophysiological recording**

Brains for recordings were obtained from adult (3-6 months) mice. After cervical dislocation (in accordance with UK Home Office license), brains were immersed in a 4°C pH 7.4 extracellular solution of 125 mM NaCl, 2.5 mM KCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 26 mM NaHCO<sub>3</sub>, 11 mM glucose and 1 mM kynurenic acid. Acute slices at 250 µm thickness were obtained using a vibrating microtome (7000smz-2, Campden Instruments Ltd., UK) and transferred to an ACSF solution oxygenated at ratio 95%:5% O<sub>2</sub>:CO<sub>2</sub> at 37°C. Slices were left to incubate for 40 minutes whilst a recording ACSF solution of 125 mM NaCl, 2.5 mM KCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 26 mM NaHCO<sub>3</sub> and 11 mM glucose was exchanged into the electrophysiological recording set up. All slice preparation and recordings were obtained in a room temperature (21°C) environment with ambient lighting conditions. Slice visualisation was achieved with infra-red illumination (M780LP1 LED, Thorlabs Ltd., UK) on two fixed stage upright microscopes (BX51W1, Olympus Ltd., Japan, and



Axioscope 2, Zeiss Ltd., Germany), both fitted with high numerical aperture objectives and digital CCD cameras (KP-M1AP, Hitachi Europe Ltd., UK). Patch pipettes were pulled from borosilicate glass capillaries (GC150F, Harvard Apparatus Inc., UK) using a two-step vertical puller (PC-10, Narishige International Ltd., UK). For voltage-clamp experiments, the internal pipette solution contained (in mM) 140 CsCl, 4 NaCl, 0.5 CaCl<sub>2</sub>, 10 HEPES, 5 EGTA, 2 Mg-ATP with pH adjusted to 7.3 using CsOH. Biocytin (1.5 mg/ml) was added to the pipette solution for post-experiment confirmation of neuronal cell type. For current-clamp recording the internal solution contained (in mM), C<sub>6</sub>H<sub>11</sub>KO<sub>7</sub> 140, NaCl 4, KCl 5, HEPES 10, EGTA 5, Mg-ATP 4, Na-ATP 0.3, CaCl<sub>2</sub> 0.5.

During patch-clamp recordings, pipette placement was controlled using micromanipulators (PatchStar, Scientifica Ltd., UK) mounted to a motorised stage (UMS, Scientifica Ltd., UK). The pipette resistances ranged between 3-4 MΩ.

For voltage-clamp experiments, the amplifier current (Axopatch 700B, Molecular Devices; Foster City, CA) was filtered at 10 kHz (−3 dB, 8-pole low-pass Bessel) and digitized at 20 kHz using a National Instruments digitization board (PCI-6052E; National Instruments, Texas, USA). Data acquisition was performed using WINWCP (Version 4.1.2) and WINEDR (Version 3.0.9) kindly provided by John Dempster (John Dempster; University of Strathclyde, UK). For current-clamp experiments, the voltage output of the amplifier was connected to a Power 1401 Series 3A (Cambridge Electronic Devices; Cambridge, UK) and was filtered at 10 kHz (−3 dB, 8-pole low-pass Bessel) before being sampled at 50 kHz.

Electrophysiological recordings of sIPSCs in the LGd and VB were obtained by Zhiwen Ye, Imperial College London. Optogenetic stimulation of the RT and mIPSC recording in LGd interneurons were obtained by Florian Zirpel, Imperial College London. Paired interneuron recordings of postsynaptic IPSCs in the LGd were obtained by Florian Zirpel, Imperial College London.

### **Monosynaptic presynaptic interneuron neuronal tracing**

Neuronal tracing of visual cortex pyramidal cell inputs to the LGd in rabies injected Sox14cre/+ mice were done using the ‘Simple Neurite Tracer’ plugin in ImageJ/Fiji (Longair et al., 2011).

Due to the large volume of coronal data required to visualise complete projections across a 2.5 mm thick volume of tissue encapsulating the visual cortex and the LGd of the thalamus, we took the following steps. A maximum intensity projection was performed across the coronal sections of tissue which covered the visual cortex pyramidal cells and the local INs in the LGd which were initially infected with the rabies virus. The maximum projection was used to crop the data to only contain signal from cortical pyramidal cells, the signal from axonal projections and dendrites, and the signal from the LGd. Axonal projections of pyramidal neurons in the visual cortex followed the lateral forebrain bundle system (lfbs) and appeared to extend anterior to the LGd before curtailing back towards the LGd. Care was taken to ensure projections were not cut off when data was cropped in either axis. Additional downsizing from  $0.54 \times 0.54 \times 10 \mu\text{m}$  to  $1 \times 1 \times 10 \mu\text{m}$  reduced the data size such that all the cropped data could be loaded in ImageJ/Fiji at once and enable tracing within a complete, intact volume. Tracing of signals belonging to dendritic and axonal structures was only done if it was evident the signal belonged to the neuron being traced. The 'Simple Neurite Tracer' algorithm uses an A-star search algorithm to automatically search for paths between two defined points (Hart et al., 1968). A significant drawback to this algorithm was the necessity to store all searches in the computer RAM. This meant that path searching in high-resolution data can be very memory intensive, causing the tracing plugin to freeze or behave erratically. To prevent software memory issues, selected points on the neurite for path searching were limited to a  $100 \mu\text{m}$  spacing in the XY coronal plane of the image, and  $10 \mu\text{m}$  spacing in the Z rostral-caudal axis. Traced neurites for neurons were skeletonized and collapsed into a single representative trace with a maximum intensity projection.

### **Assessing specificity of targeting LGd**

To assess the targeting of the LGd during rabies tracing, free-floating slices containing the LGd were extracted (n=28, one rabies sample) from the cooled PBS and mounted using glycerol mounting medium and stored at  $4^{\circ}\text{C}$ . Sections were imaged under a Nikon Eclipse 80i inverted widefield microscope using a  $10\times$  objective and emission collected across two bandpass filters, FITC 535/45 nm and Texas Red 630/60 nm, for GFP and mCherry fluorescence collection respectively. The number of green (presynaptic and rabies-infected), yellow (rabies-infected starter cells) and red (starter cells) were manually counted across the LG.

### **Interneuron pair immunolabelling**

Biocytin (2 mg/ml) added to the internal pipette solution was used to fill interneuron pairs during whole-cell patch-clamp recording. Each 250  $\mu\text{m}$  thick slice was fixed in 4% PFA for 24 hours. The following day slices were washed three times for 10 minutes each using 4°C PBS. Slices were then treated with 10% goat serum and 0.5% Triton-X for 3-4 hours at room temperature. After an additional wash with 4°C PBS for 10 minutes, slices were treated with 1% goat serum, 0.5% Triton-X and 1:1000 anti-rabbit GFP for 24 hours at 4°C. Slices were washed four times for 10 minutes each before treating with 1% goat serum, 0.5% Triton-X and 1:200 Alexa Fluor 555 streptavidin for 3 hours. Slices were washed three more times with 4°C PBS for 10 minutes each, before mounting with fluorescence mounting medium (Vectashield). The immunolabelling procedure was conducted by Florian Zirpel, Imperial College London.

### **Imaging of interneurons**

A Leica SP8 confocal microscope system equipped with 40 $\times$  NA 1.3 mm glycerol immersion objective was used to image tissue sections containing biocytin filled interneuron pairs. To cover the complete arborisation of the interneurons within the slice, 2  $\times$  2 tiling was required in XY at 3000  $\times$  3000 pixels per tile (covering a physical length 387.5  $\mu\text{m}$ ), with a spatial resolution of 0.129  $\times$  0.129  $\mu\text{m}$ . Z-stacks were generated over 200-250  $\mu\text{m}$  with a Z-resolution of 0.339  $\mu\text{m}$ . A DPSS 561 nm laser was used for Alexa 555 excitation and emission collected over a single PMT set to 568-688 nm. The laser power was adjusted over the sample depth to compensate for laser attenuation from scattering. Imaging took over 6 hours to complete and generated around 50 GB of image data. Tiled sections were stitched using the Grid/Collection stitching plugin in ImageJ/Fiji using a manually adjusted overlap to generate seamless alignment.

### **Interneuron reconstruction**

The intricate morphology of interneurons spanned over 500  $\times$  500  $\mu\text{m}$  in XY. This meant that manual tracing was unfeasible for the following reasons. The complete 50 GB data set would need to be loaded simultaneously to ensure processes from both interneurons could be distinguished from themselves and each other. In addition, interneuron processes spanned across the complete Z-stack comprised of around 700 slices. Manual tracing efforts,

therefore, required simultaneous tracking of many individual processes whilst only being able to sample a single slice at a time. Paired interneuron fills also showed multiple close appositions making it impossible to distinguish which neurites belonged to any interneuron. The high-resolution afforded by EM imaging as in Morgan & Lichtman (2020) allowed for easier distinction between individual neurites, albeit at the sacrifice of a much larger data set.

To address the challenge of tracing complex morphology, we used a deep-learning reconstruction approach. The same U-Net architecture and augmentation pipeline developed in 'Chapter 3' was used to segment the paired interneuron image data, with a few modifications to the training regime, as follows. Training data was generated from 41 randomly sampled  $512 \times 512$  pixel regions, with all samples presenting both neurite signal and background signal. This was different from the automated cell detection pipeline where some samples contained only background signal. Neurite morphology was also significantly more complex than soma morphology, which is spherical-like in geometry. Neurites can appear as long winding fibres running transverse to the imaging plane, with short high-intensity locality when running orthogonal to the imaging plane, or anywhere in between.

The image data contains signal from both the in-focal plane and out-of-focal plane. Therefore, to ensure the network was not sensitive to detecting patterns from either the out-of-focal plane signal or background autofluorescence, manual segmentation was done using the criteria that manually labelled fluorescent signal had to be high intensity, as if the structure was being directly sampled from the in-focal plane. Any lower intensity signal was omitted even if they clearly corresponded to neurite signal. The ground-truth data set, therefore, reflected regions where neurites were likely to be sampled in the image plane with high intensity, and credence was given to the network for segmenting the lower intensity neurite structures only when they appeared with high intensity at a subsequent Z-stack position. After following the same augmentation pipeline outlined in 'Chapter 2', the network was trained using the ELU activation function, a BCE loss, an Adam optimiser initialised with a learning rate of  $1e-4$ , a batch size of 8 for feeding the network with training samples, and validation loss was evaluated using the Dice loss score. The training was executed on an Ubuntu workstation with an Intel(R) Core(TM) i9-7980XE CPU at 2.60GHz with 32 cores and 128 GB RAM, equipped with an NVIDIA Titan Xp GPU.

To reduce false-positive predictions when using the trained model, incremental maximum intensity projections of 100 consecutive slices were generated and traced with a brush in ImageJ/Fiji to create a mask. Each projection mask was combined to produce a single mask that could be projected across each slice in the confocal image stack. Incremental maximum intensity projections were chosen over the use of a single projection across the Z-stack, to ensure low-intensity neurite structures were not obscured by high-intensity noise elsewhere in the stack. For example, image noise was typically higher towards the top of the Z-stacks, due to surface distortions and reduced laser attenuation. The mask was applied to each slice in the confocal Z-stack and split into  $512 \times 512$  pixel blocks for parallel prediction with the U-Net network. Predicted blocks were thresholded at 0.5 to binarise the prediction into neurite signal class (value 1) and background signal class (value 0). Each slice was then reconstructed from the predicted  $512 \times 512$  blocks. Prediction of the 50 GB data sets took one hour to complete and resulted in another 50 GB data set containing predicted segmentation of the confocal data sets.

A data set this significant in size was not practical to view in any software, so the segmented data set was converted into a mesh by using a cube marching algorithm with a step size of 2 pixels (Lewiner et al., 2003). This process iterated a  $2 \times 2 \times 2$  cube kernel over the data, creating vertices and edges wherever the binary segmentation had a bisection with an edge of the cube kernel. An input data set of  $6000 \times 6000$  pixels and 700 slices totalled to 25 trillion voxels and resulted in a mesh containing around 18 million vertices following the cube marching process. The mesh was saved as a stereolithography (STL) file, then opened in the open-source 3D computer graphics software, Blender, and decimated down to 500-600 thousand vertices. This resulted in a 15 MB mesh for each interneuron pair (reduced from 50 GB), which could be loaded on any workstation or online web tool. Each mesh could be viewed in three-dimensions at the detail of tracking individual neurite processes.

### **Analysing dendritic morphology**

Dendritic diameters were stochastically measured a maximum projection in XY using the confocal data sets for each interneuron pair, using Fiji/ImageJ. Visual identification of thicker and thinner dendrites based off dendritic branch order was used to manually classify

measures into thicker trunk-like or thinner targeting dendrites for diameter statistical measures.

To determine whether the underlying distribution of all (unclassified) diameter measurements could be best described by two classifications of thick and thin dendritic morphology, as manually chosen above, Gaussian mixture modelling was used, and implemented through SciPy (Pedregosa et al., 2011). For an unsupervised selection of the optimum number of Gaussian mixtures (clusters), the model with the lowest Bayesian information criterion (BIC) was chosen (Schwarz, 1978).

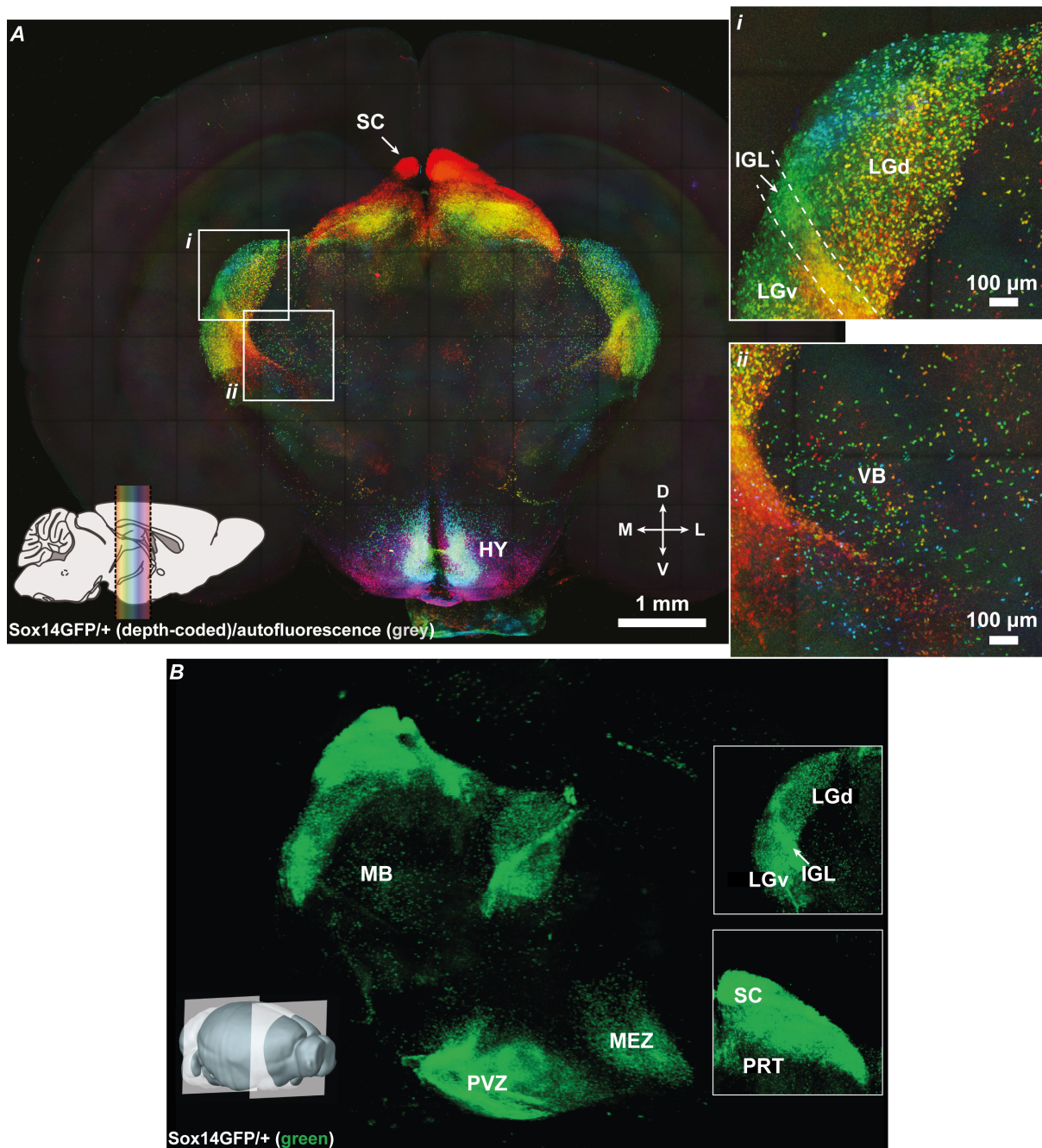
### **Putative anatomical interneuron-interneuron connectivity**

Interneuron pair meshes were plotted using the Plotly library in combination with a custom dashboard web application (Dash framework). Plotly is more suited to large data set plotting in comparison to the Matplotlib plotting library, hence its use here. The Dash framework was used to create an application which plotted interneuron-interneuron meshes alongside a coordinate input field text-field. The mesh could then be manipulated in three-dimensions, and close appositions ( $<10\ \mu\text{m}$ ) were marked by entering their coordinate position into the web app. Appositions and putative anatomical connections were confirmed by exporting the marked coordinate positions and locating each position in the original high-resolution confocal data set. We then validated any close appositions by visually assessing the confocal data set using the marked coordinate positions.

### 5.3 RESULTS

#### **Sox14 interneurons widely distributed across the thalamus**

STP imaging of endogenous GFP+ fluorescence in the Sox14gfp/+ mouse showed a widespread distribution of Sox14 expressing cells across the entire thalamus. Previous studies have reported an association between the Sox14 gene with GABAergic neurons in the diencephalon during embryonic development (Delogu et al., 2012). However, the extent of Sox14 presence outside of the visual thalamus relay nuclei has only been reported once before (Jager et al., 2016), and here it can be seen that Sox14 cells are distributed across the entire thalamus (Fig. 5.1). Dense Sox14 gene expression can be observed in the LGd, previously reported to colocalise with the local GABAergic interneuron population (Jager et al., 2016), as well as the LGv and IGL. Dense GFP+ neurons were also observed in the sensory and motor-related superior colliculus (SC) of the midbrain region, the nociceptive relay nuclei of the ventrobasal complex (VB), and numerous domains of the hypothalamus including the periventricular zone (PVZ) and the medial zone (MEZ). This observation qualitatively suggests a larger population of interneurons outside of the visual thalamus than previously thought.



**Figure 5.1.** STP imaging across the thalamus of a *Sox14gfp/+* mouse. **A.** GFP positive cell expression is widely distributed across thalamic, hypothalamic and midbrain nuclei. Extremely high expression levels can be observed in SC making single-cell detection impossible. High GFP expression can be observed in LGd, IGL and LGv regions as expected. LGd *Sox14* cells are also known to be local interneurons enabling effective detection of interneuron population. The ventrobasal complex showed sparse expression. **B.** STP imaging at  $0.54 \times 0.54 \times 10 \mu\text{m}$  permits three-dimensional modelling by thresholding high GFP signal from background. The structural formation of hypothalamic nuclei such as the PVZ and MEZ, can be more easily identified. Superior colliculus, SC; dorsal lateral geniculate nucleus, LGd, intergeniculate leaflet, IGL,



*ventral lateral geniculate nucleus, LGv, hypothalamus, HY; midbrain, MB; hypothalamic periventricular zone, PVZ; hypothalamic medial zone, MEZ; pretectal region, PRT.*

### **Sox14gfp/gfp (knockout) shows significantly lower Sox14 gene expression**

GFP+ Sox14 neuron distribution in the Sox14gfp/gfp (knockout) model was markedly reduced. In particular, the SC, midbrain reticular nuclei (MRN) and visual thalamus with the LGd and LGv showed much lower densities (Fig. 5.2). This result agrees with previous reports that local interneurons migrate from a prethalamus region during the first postnatal week and that Sox14 ablation restricted cell differentiation and therefore the migratory event (Golding et al., 2014).

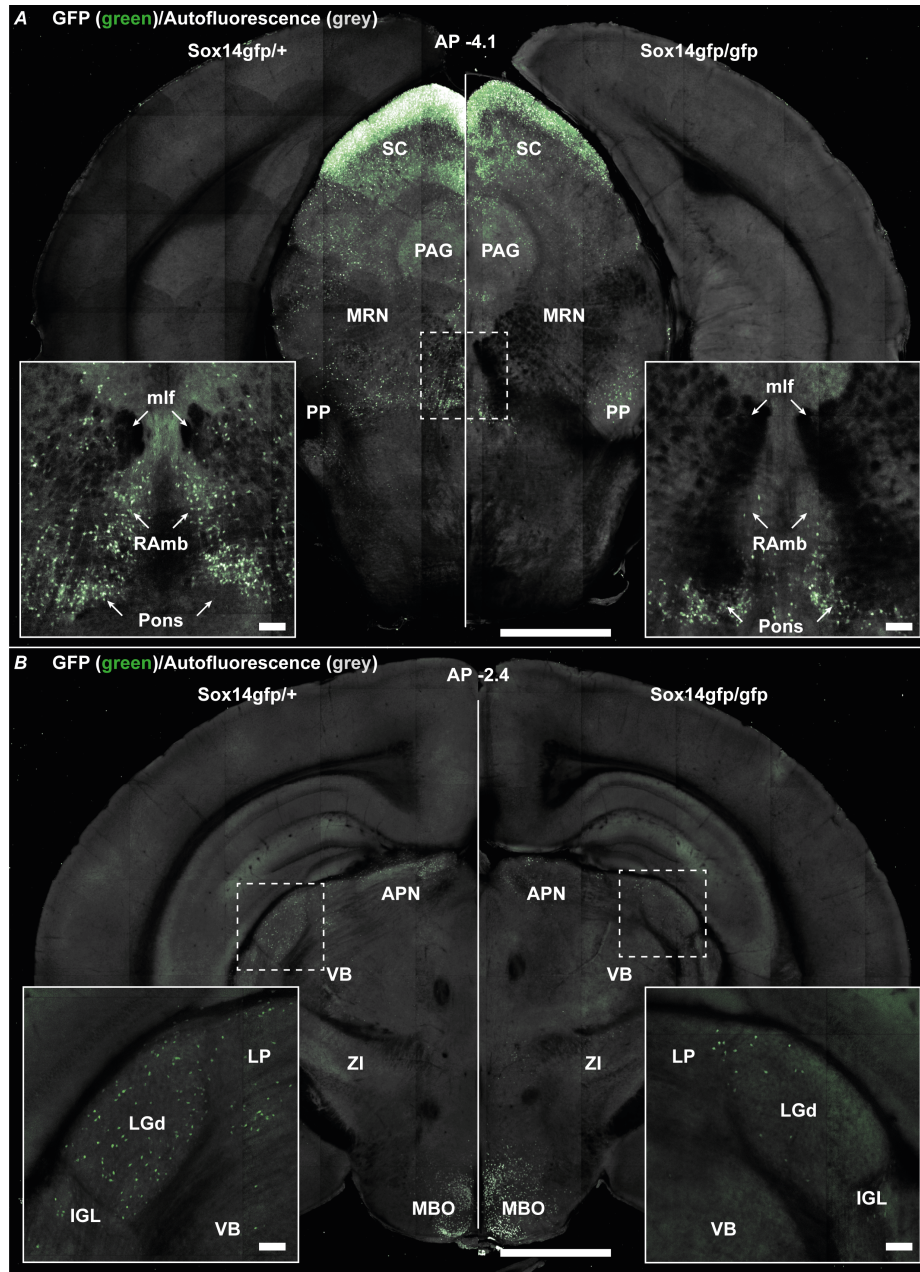
### **Higher prevalence of Sox14 interneurons in first-order sensory nuclei followed by higher-order nuclei**

To quantitatively evaluate Sox14 gene expression in the thalamus, automated cell detection using a trained U-Net following CCF3 atlas registration was performed, as described in methods of 'Chapter 4', on the Sox14gfp/+ (P21, n=3 mice) and Sox14gfp/gfp (P21, n=2 mice) mouse. In association with the validation study performed in 'Chapter 4', the number of GFP+ cells in the LGd was automatically counted at  $1,234 \pm 82$  (mean  $\pm$  SD, n=3 mice), in comparison to a stereological study which reported  $1,255 \pm 195$  (mean  $\pm$  SD)(Evangelio et al., 2018), showing remarkable agreement (Fig. 5.3A).

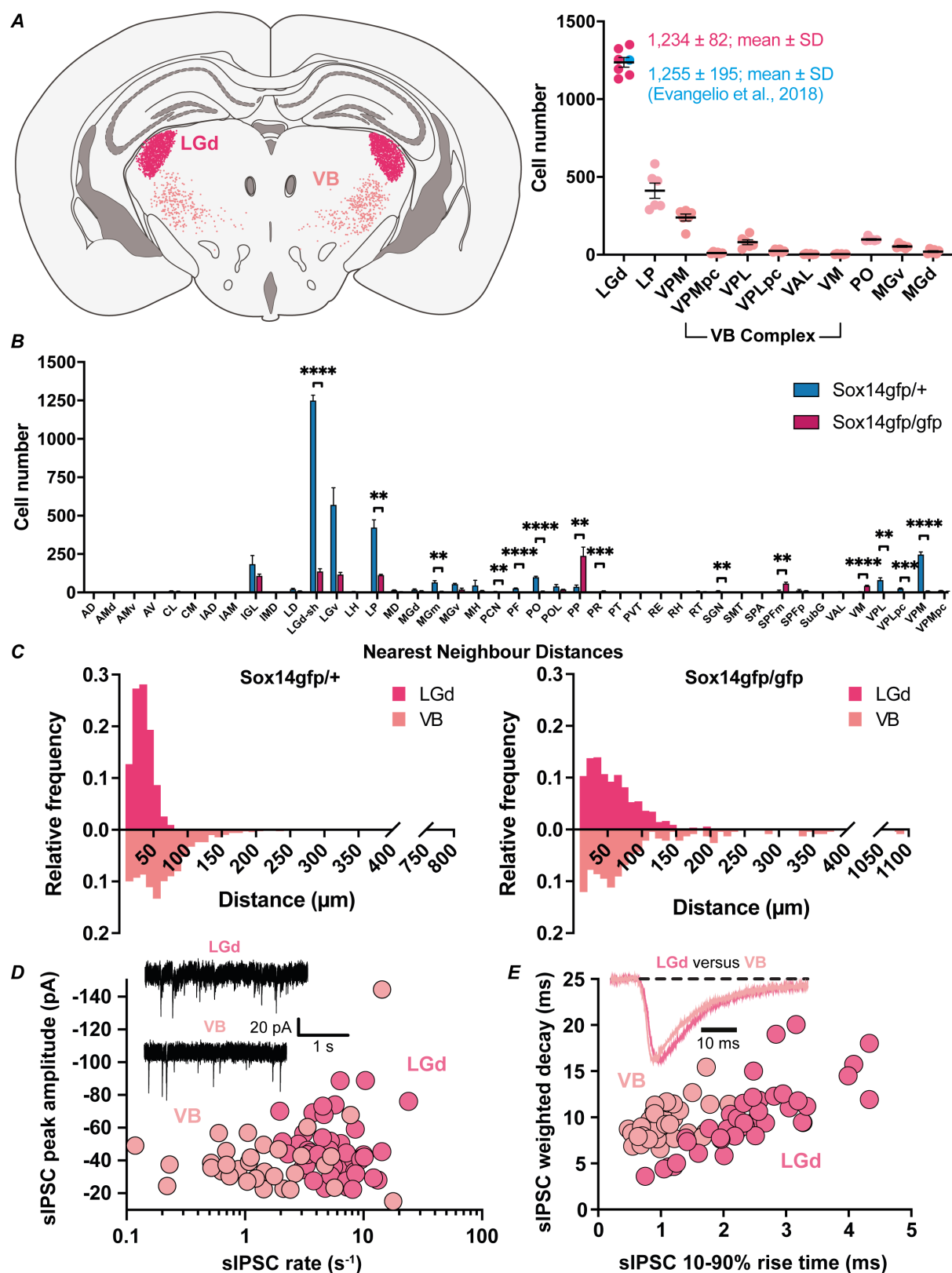
In the Sox14gfp/+ mouse, a total of  $6,588 \pm 811$  Sox14 cells (mean  $\pm$  SD, n=3 mice) were counted across all nuclei in the thalamus across both hemispheres. Higher GFP+ density was observed in sensory and motor-related regions of the thalamus with increased prevalence in first-order nuclei. In the visual region of the thalamus, first-order LGd had  $1,234 \pm 82$  GFP+ cells, and higher-order LP had  $411 \pm 110$  cells. The sensory-motor VB complex divided into ventroposterior medial nucleus (VPM) had  $250 \pm 49$  cells, the ventroposterior lateral nucleus (VPL) had  $105 \pm 40$  cells, the ventral anterior lateral nucleus (VAL) had  $4 \pm 3$  cells, and the ventromedial nucleus (VM) had  $4 \pm 2$  cells. The higher-order posterior nucleus (PO) had  $99 \pm 12$  cells.

In the auditory region of the thalamus, the first-order ventral medial geniculate (MGv) had  $53 \pm 12$  cells, and higher-order dorsal medial geniculate (MGd) had  $20 \pm 13$  cells. Sparse GFP+ detection was also observed in associative TC nuclei such as the laterodorsal nucleus (LD) with  $20 \pm 7$  cells, the parafascicular nucleus (PF) with  $25 \pm 5$  cells and mediodorsal nucleus (MD) with  $13 \pm 6$  cells. Counts represent a single hemisphere with averaging across both hemispheres and all samples.

Unpaired, two-tailed t-Tests, assuming non-consistent standard deviation, on the Sox14gfp/+ (n=3 samples with two hemisphere measures per sample) mouse against the Sox14gfp/gfp (knockout, n=2 samples with two hemisphere measures per sample) showed a significant reduction in GFP+ cells in the LGd (listed as LGd-sh in Fig. 5.3B,  $p < 0.000001$ ), posterior complex (PO,  $p = 0.000001$ ), central posteromedial nucleus (VPM,  $p = 0.000004$ ), ventral medial nucleus (VM,  $p = 0.000004$ ), parafascicular nucleus (PF,  $p = 0.000013$ ), parvicellular ventral posterolateral nucleus (VPLpc,  $p = 0.000341$ ), perireunensis nucleus (PR,  $p = 0.000794$ ), lateral posterior nucleus (LP,  $p = 0.001229$ ), parvicellular subparafascicular nucleus (SPFm,  $p = 0.001675$ ), paracentral nucleus (PCN,  $p = 0.001713$ ), medial region of the medial geniculate complex (MGm,  $p = 0.001713$ ), ventral posterolateral nucleus (VPL,  $p = 0.002583$ ), peripeduncular nucleus (PP,  $p = 0.003207$ ), and the supragenulate nucleus (SGN,  $p = 0.003501$ ).



**Figure 5.2.** Qualitative comparison of Sox14gfp/+ and Sox14gfp/gfp mice at AP -4.1 (A) and AP -2.4 (B) with enlarged regions indicated by dotted lines. Sox14gfp/gfp (knockout) shows decreased GFP expression across midbrain, thalamic and hypothalamic regions. In particular, GFP expressing cells in the LGd were lower owing to a reduction in interneuron number. Superior colliculus, SC; periaqueductal grey, PAG; midbrain reticular nucleus, MRN; peripeduncular nucleus, PP; medial longitudinal fascicle, mlf; midbrain raphe nuclei, RAmb; anterior pretectal nucleus, APN; ventrobasal complex, VB; zona-incerta, ZI; lateral posterior nucleus, LP; dorsal lateral geniculate nucleus, LGd, intergeniculate leaflet, IGL; mammillary body, MBO. Scale bars 1 mm in overview panels, 100 μm in enlarged panels.



**Figure 5.3.** Quantitative mapping of GFP+ neurons in Sox14gfp/+ and Sox14gfp/gfp mice. **A.** Exemplar distribution of GFP+ neurons in the LGd and VB (left) alongside automated cell detection across thalamic nuclei (right). Validation of automated cell detection is included as  $1,234 \pm 82$  (mean  $\pm$  SD,  $n=3$ ) automatically detected GFP expressing cells in the LGd compared

to  $1,255 \pm 195$  from Evangelio et al. (2018). **B.** Automated cell detection across the entire thalamus in both the Sox14gfp/+ (blue,  $n=3$  samples with two measures per sample) and Sox14gfp/gfp (red,  $n=4$  samples with two measures per sample) mouse. Significant differences from unpaired, two-tailed t-Tests are indicated as \*\* ( $p<0.01$ ), \*\*\* ( $p<0.001$ ), \*\*\*\* ( $p<0.0001$ ), with measurements taken for each hemisphere, not assuming consistent standard deviation. LGd is represented as LGd-sh. **C.** Nearest neighbour distances between interneurons of the LGd and VB complex in the Sox14gfp/+ ( $n=3$ ) and Sox14gfp/gfp ( $n=2$ ) mice. **D.** sIPSC peak amplitude against instantaneous IPSC rate recorded from relay cells in the LGd and VB. sIPSCs were automatically detected from current traces (black). **E.** sIPSC weighted decay against the 10-90% rise time shows longer rise times in LGd relay cells, compared to the VB. Average sIPSC waveforms for the LGd and VB illustrated in plot, aligned to the initial rising phase.

### **Distinct spatial distributions of Sox14 neurons within LGd and VB complex**

From automated cell detection, we calculated the distance to the nearest neighbour for both the Sox14gfp/+ and Sox14gfp/gfp mouse (Fig. 5.3C). Distinct differences in nearest neighbour distances were observed between the LGd and VB complex in the Sox14gfp/+ mouse with a modal distance of  $35 \pm 5 \mu\text{m}$  (mean  $\pm$  SD,  $n=3$  mice) in the LGd compared to  $55 \pm 5 \mu\text{m}$  in the VB complex. A similar distribution was observed in the knockout model, although reduced GFP+ presence in the LGd increased the range of nearest neighbour distances (Fig. 5.3C). To functionally classify this observed difference in distribution, whole-cell voltage-clamp recordings of spontaneous inhibitory postsynaptic currents (sIPSCs) were made in relay cells of the LGd and the VB complex. An average of  $460 \pm 64$  (mean  $\pm$  SD) sIPSCs was detected in 43 cells recorded from the LGd with a similar number of sIPSCs ( $403 \pm 66$  events) detected in 32 cells recorded from the VB. The magnitude of the synaptic conductance was similar in the LGd ( $-47.6 \pm 2.6 \text{ pA}$ ) compared to the VB ( $-40.6 \pm 4.0 \text{ pA}$ ). However, the mean instantaneous firing rate of LGd relay neuron sIPSCs was  $6.5 \pm 0.6 \text{ s}^{-1}$  compared to the VB, which was significantly lower at  $2.7 \pm 0.7 \text{ s}^{-1}$  (unpaired, two-tailed t-Test,  $p=0.0001$ ). The LGd relay neurons observed reduced variability in sIPSC instantaneous rates compared to the relay neurons in the VB (Fig. 5.3D).

To address if the difference in the distribution of local interneurons in the LGd and VB impacted upon the sIPSC waveforms, the average waveform was constructed and evaluated for 10-90% rise time and decay (Fig. 5.3E). The decay was not significantly different (unpaired, two-tailed t-Test,  $p=0.49$ ) between the relay neurons of the LGd at  $9.93 \pm 0.66$  ms (measures from 43 cells) and the VB at  $9.44 \pm 0.3$  ms (measures from 32 cells). However, the 10-90% rise time for the LGd was significantly slower (unpaired, two-tailed t-Test,  $p<0.005$ ) at  $2.42 \pm 0.13$  ms compared to the VB at  $1.08 \pm 0.07$  ms, suggesting closer proximity of interneurons leads to longer 10-90% rise time, but does not influence the decay.

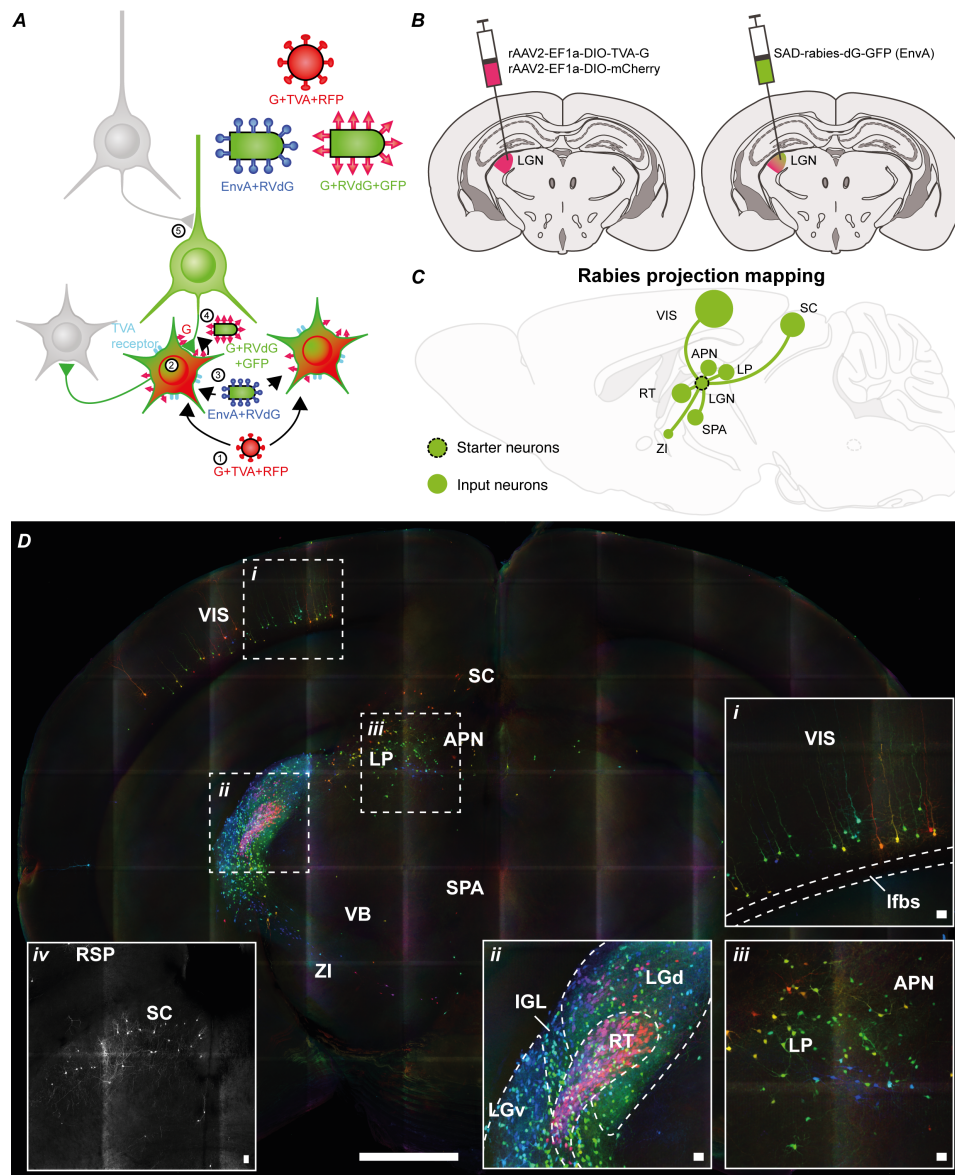
### **Sox14 neurons in LGd receive monosynaptic input from both cortical and thalamic nuclei**

To map monosynaptic inputs to interneurons of the LGd, a modified G-deleted rabies virus was injected into the LGd of the Sox14cre/+ mouse. Observed afferent inputs to Sox14 neurons in the LGd primarily came from VIS, APN, LP, SPA, ZI and unexpectedly, the RT and SC (Fig. 5.4). The RT is known to provide GABAergic drive primarily to thalamocortical neurons in the dorsal thalamus and assists in the generation of sleep rhythms (Steriade, 2005). However, its anatomical connectivity to local interneurons in the LGd has not been observed or functionally reported. RT inputs to LGd interneurons appeared to be confined to the most dorsal region of the nucleus, which was directly anterior to the LGd, suggesting a topographic arrangement of projection neurons. Lack of Sox14 cells in the RT (Fig. 5.3B) confirms that these cells were labelled through the transsynaptic spread of infected Sox14 cells in the LGd. However, the efficacy of specifically targeting the LGd was problematic (discussed later). There was also evidence of projections from the superficial layer of the SC. Pathways between the SC and LGd have only been previously shown to exist with the LGd TC relay cells and has not been shown to exist in the interneuron population until now (Bickford et al., 2015).

Classification of labelled neurons in the visual cortex showed most inputs to the LGd originated from layer 5 and 6 of primary visual cortex (VISp), with  $1.70\% \pm 0.72\%$  (mean  $\pm$  SD,  $n=3$  mice) pyramidal cells projecting from layer 5 VISp,  $4.82\% \pm 0.68\%$  from layer 6a, and  $0.36\% \pm 0.25\%$  from layer 6b (Fig. 5.5A). Percentages are with respect to the total number of detections across all visual cortex areas. The lateral visual area (VISl) showed the next highest input with  $0.27\% \pm 0.33\%$  inputs originating from layer 5,  $0.72\% \pm 0.29\%$  from layer 6a and  $0.02\% \pm 0.04\%$  from layer 6b. Very sparse projections also originated from the ipsilateral



anterolateral visual area (VISal) and anteromedial visual area (VISam). Further inputs were seen originating from layer 1 visual cortex. However, these abnormally high detections resulted from autofluorescence noise detected as cells on the surface of the cortex. This reveals an apparent limitation in the method of automated detection of cells.

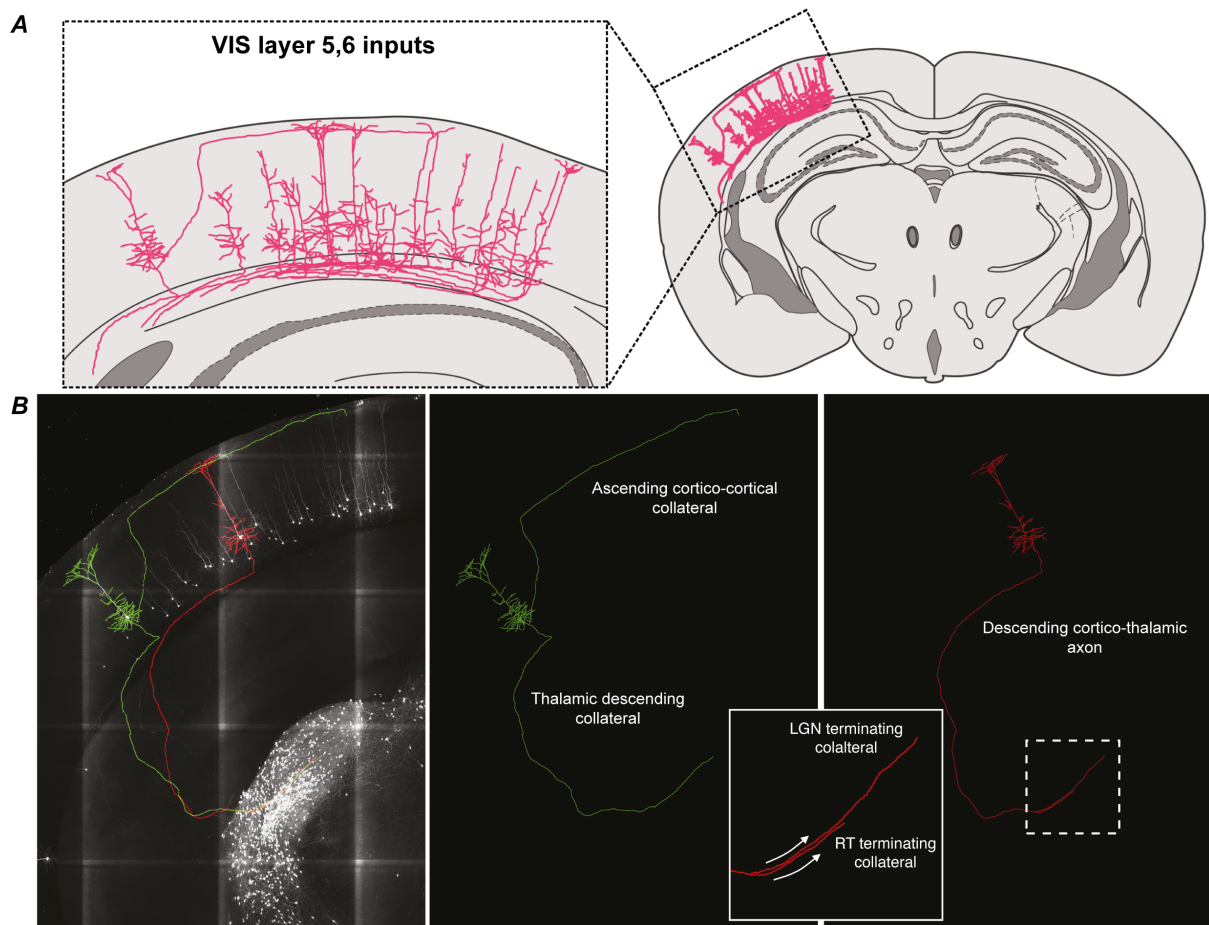


**Figure 5.4.** Presynaptic mapping to Sox14 expressing cells in the LGd. **A.** Schematic of rabies infection. 1) Cells are primed for rabies infection by injecting an AAV containing G-protein and TVA into the Sox14 neuron population. 2) Primed cells express G-protein and TVA receptors. 3) G-deleted rabies virus expressing EnvA protein is injected into the same region and is internalised by primed cells following EnvA binding to TVA receptors. 4) Primed and rabies-infected cells produce G-deleted rabies virus which binds with G-protein in primed cell and can

*infect neurons presynaptically. 5) Lack of G-protein in presynaptically infected neurons prevents the rabies virus from further presynaptic spread, limiting monosynaptic labelling of cells presynaptic to the initially infected cells. B. Injection target LGd for rabies tracing. C. Mapping of rabies inputs to infected starter cell population following rabies infection. D. Depth-coded intensity projection across 2 mm thick coronal volume following rabies tracing. Projections from VIS (i) form a large number of inputs, with sparse input mapping from LP, APN (ii) and ZI. RT (iii) also showed projection neurons suggesting inhibitory input to local interneurons in the LGd. Projections from SC (iv) were also identified (section is taken from another slice projection). Visual cortex, VIS; anterior pretectal nucleus, APN; ventrobasal complex, VB; zona-incerta, ZI; lateral posterior nucleus, LP; dorsal lateral geniculate nucleus, LGd, intergeniculate leaflet, IGL, ventral lateral geniculate nucleus, LGv; reticular nucleus, RT; lateral forebrain bundle system, lfbs; subparafascicular nucleus, SPA; superior colliculus, SC. Scale bar 1 mm in overview and 100  $\mu$ m in enlarged panels.*

High-resolution near-complete neuronal tracing of two exemplar pyramidal neurons from layer 6 of VISp showed two distinct axonal morphologies. One pyramidal neuron (green)(Fig. 5.5B) showed dendritic branching up to the superficial layers of the cortex as expected, with an axon projecting towards the lateral and medial forebrain bundle system. However, after forming from the soma, the axon split into two collaterals. One neuron axon projected back into the superficial layers of VISp and travelled medially before terminating and presumably innervating proximal dendrites. The second neuron axon continued to project towards the bundle systems where it travelled towards the LG until the axon was lost amongst other neurite structures, just anterior to the LGd. This pyramidal neuron presented both corticocortical and corticothalamic projections. The second pyramidal neuron (red) presented similar dendritic morphology to the first, with an axon which projected through the bundle system. At a point anterior to the LG and RT, the axon split with one collateral appearing to terminate in the RT whilst the other continued to the LG where it is lost amongst signals from other neurites. The axonal projection targeting the RT is possibly required for the indirect recruitment of GABAergic inhibition onto interneurons in the LG. The other axonal projection then directly synapses onto the interneurons. The RT, therefore, appears to play an important for both thalamocortical corticothalamic pathways with TC innervation of the RT and now evidence of corticothalamic innervation of the RT en-route to the LGd.



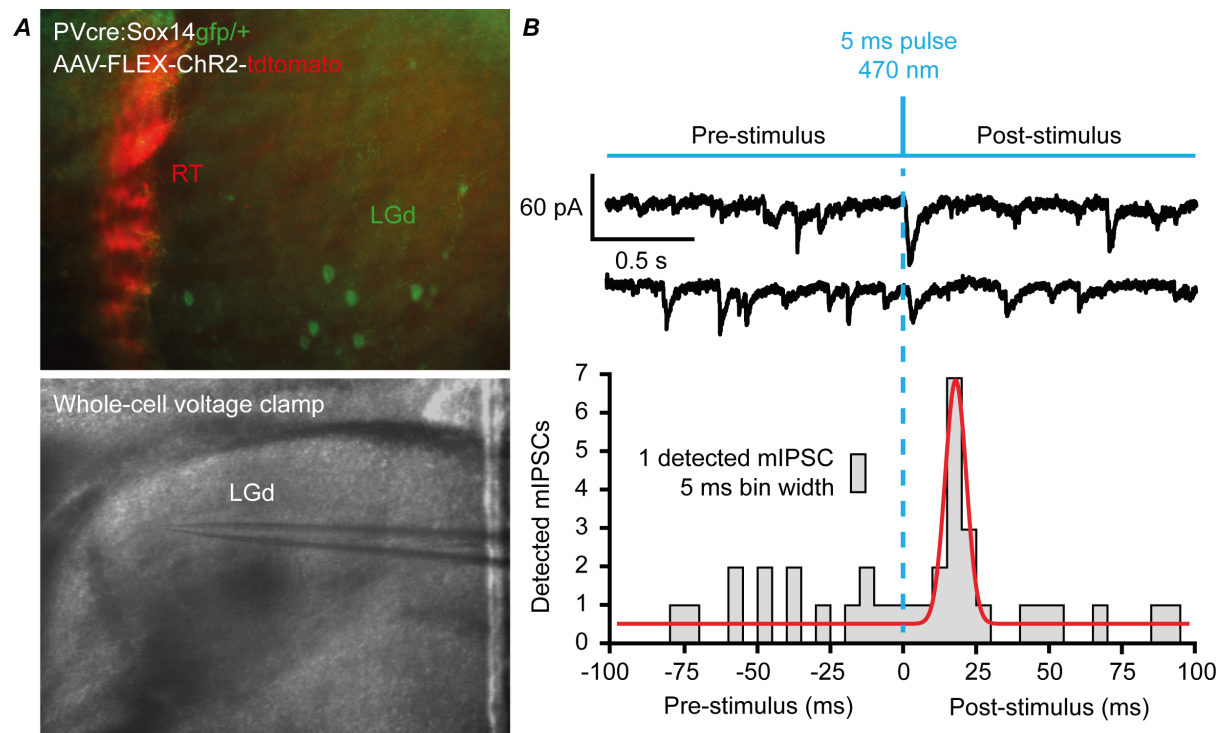


**Figure 5.5.** Mapping visual cortex inputs to LG. **A.** Exemplar neuronal tracing of layer 5 and 6 pyramidal neurons showing projections flowing along the forebrain bundle system and towards the rostral thalamus. **B.** Near-complete tracing of two pyramidal neurons shows both axon collaterals. One pyramidal neuron (green) shows cortico-cortical axon travelling into the superficial layers of the visual cortex and projecting medially, before terminating — the other collateral projects towards the RT. The second pyramidal neuron (red) shows a single axon projecting from the visual cortex down towards the RT where it splits to send another axon towards the LG before it is lost.

### RT shows functional input to LGd interneurons

To check whether the observed anatomical connectivity from RT GABAergic neurons to LGd interneurons from the monosynaptic rabies tracing was functionally present, a PVcre;Sox14gfp/+ crossed transgenic mouse strain was injected with AAV-FLEX-ChR2-tdtomato (Fig. 5.6A). From the rabies tracing result, the injection site was targeted to the

most dorsal region of the RT, which was positioned directly anterior to the LGd, where there appeared to be the highest density of GABAergic inputs. Following channelrhodopsin (ChR2) expression via Cre-mediated recombination in GABAergic projection neurons of the RT, whole-cell voltage-clamp recordings of LGd Sox14 interneurons were performed during optogenetic stimulation of the RT with 5 ms pulses at 470 nm. A high rate of spontaneously occurring mini inhibitory postsynaptic currents (mIPSCs) was present in all recordings meaning a quantitative analysis was needed to identify mIPSCs that were correlated with the RT stimulation. A peristimulus time histogram (PSTH) of detected mIPSCs was constructed with a 5 ms bin width that showed the occurrence of mIPSCs at a latency of 20 ms following the optogenetic stimulus (Fig. 5.6B). Stimulus evoked mIPSC activity provided evidence for functional connectivity between RT neurons and interneurons of the LGd in 7 out of 33 Sox14 interneuron recordings. Unconstrained Gaussian fits to the PSTH distributions obtained from these 7 cells gave  $R^2$  values ranging from 0.70 to 0.99 ( $0.84 \pm 0.05$ ,  $n=7$ ) with a central peak at  $17.2 \pm 1.52$  ms and a width of  $8.41 \pm 1.75$  ms.



**Figure 5.6.** Whole-cell voltage-clamp recordings of Sox14 interneurons of the LGd following stimulation of RT. **A.** PVcre;Sox14<sup>gfp/+</sup> transgenic mouse was injected into the most dorsal region of the RT with AAV-FLEX-ChR2-tdtomato. Following flexed-switched expression of ChR2 in RT GABAergic neurons, whole-cell voltage-clamp was performed on Sox14 interneurons in

the LGd. **B.** Peristimulus time histogram with a bin width of 5 ms describing the occurrence of detected mIPSC events that were correlated with the optogenetic stimulation of the RT with 470 nm, 5 ms light pulses. Note the presence of asynchronous mIPSCs that were not correlated with the optogenetic stimulus.

### **Rabies targeting of LGd also hits IGL and LGv interneurons**

Although rabies injections were targeted to LGd, specific infection of local interneurons in the LGd proved a challenging task. The LG, consisting of LGv, IGL and LGd show extensive Sox14 expression, and therefore all closely located nuclei were prime candidates for rabies infection (Fig. 5.1). In all samples, there was evidence of infection within the LGd, IGL and LGv with green and red fluorescence outside the LGd. In one sample, of a total of 642 primed cells, rabies infection of primed cells only occurred in the LGv and IGL region (18 cells), with presynaptic labelling (green fluorescence only) observed across LG and RT (total 388 cells). This corresponds to approximate labelling of 22 presynaptic cells per successful rabies-infected primed cell. Monosynaptic tracing is therefore reflective of general Sox14 neuron connectivity of the LG and not specifically interneurons in the LGd as intended. Dense labelling and the ineffective distinction between red and green fluorescence with the STP tomography setup prevented the specific mapping of inputs to the LGd. Modifications to the injection protocol (not shown here) including in-utero rabies infection in the LGd did not show any benefit. Anatomical tracing with the rabies virus was, therefore, limited but the data did identify the possibility that LGd interneurons receive the synaptic drive from the surrounding brain regions such as the RT for which the existence of this previously unreported connectivity was verified using optogenetic methods.

### **LGd interneurons are morphologically complex**

Using the same U-Net architecture (see 'Chapter 4') trained on manually segmented neurite samples, 50 GB confocal Z-stacks of paired interneurons biocytin fills were automatically segmented (Fig. 5.7). Compared to a thresholding approach using Otsu's algorithm, the deep-learning approach generated superior detection of neurite signal with fewer false-positive detections (Fig. 5.7ii). The tubular neurite structures were more accurately segmented and complete compared to Otsu's thresholding method (Fig. 5.7iii). Reduced false-positive detection was necessary to reduce the overall size reconstructed neuron when converted into

a mesh. The maximum intensity projection mapping across 100 sequential slices also assisted in suppressing segmentation of noise signal in large volumes of the data set where no neurite signal was present.

Applying a cube-marching algorithm to the predicted segmented output was necessary for converting the 50 GB prediction data set to a more manageable 15 MB mesh which could be loaded into three-dimensional graphics software such as Blender, or even online viewers. Plotting the mesh with Plotly and Dash (Fig. 5.8A), enabled close appositions ( $<5\ \mu\text{m}$ ) between paired interneurons to be visually identified and marked. Coordinate positions were exported and used to extract volumes from the original confocal data set to validate the apposition (Fig. 5.8A).

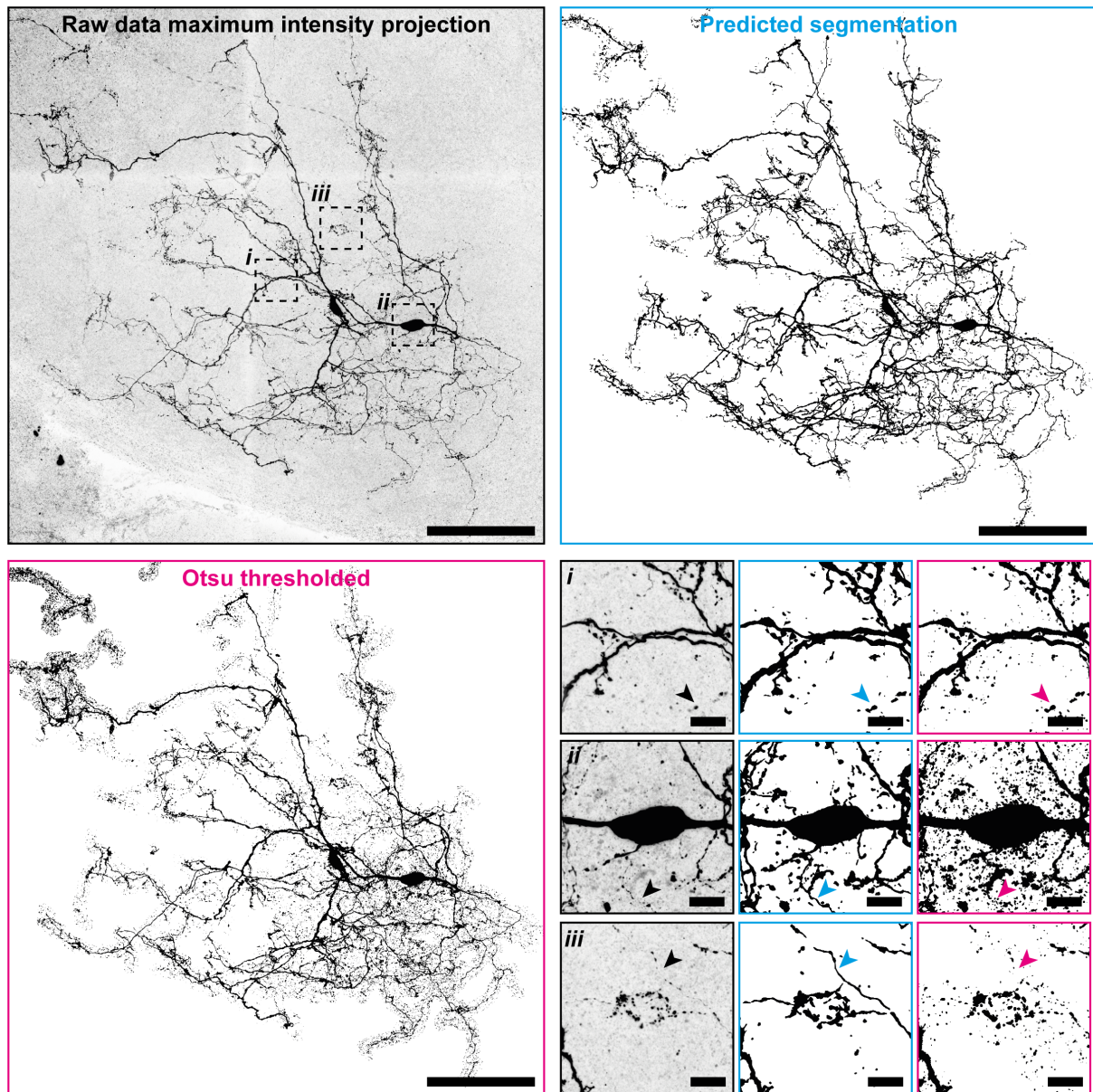
Individual interneurons showed exceptional morphological complexity, spanning across a 500-600  $\mu\text{m}$  diameter in XY (coronal plane) and across the entire 250  $\mu\text{m}$  Z-depth (rostral-caudal plane) of the slice preparation (Supplementary Fig. 5.1). This dendritic spanning was in agreement with the comprehensive single interneuron reconstruction recently published by Morgan & Lichtman (2020). Thicker trunks originating from the soma and circuitous targeting dendrites were distinguishable by their morphology. Whilst axons were likely to arborise from the trunks, inability to map RGC terminals, or lack thereof, meant the axon could not be identified by morphology in the confocal data. Trunk dendrites were measured at  $1.49 \pm 0.74\ \mu\text{m}$  (mean  $\pm$  SD;  $n=189$  measures across 6 interneuron pairs), which were almost three times the size as the targeting dendrites at  $0.52 \pm 0.19\ \mu\text{m}$  ( $n=581$  measures across 6 interneuron pairs). Unpaired t-test with Welch's correction for non-equal standard deviation showed significantly different classification of the trunk and targeting dendrite diameters ( $p<0.0001$ ). In comparison, Morgan & Lichtman (2020) measured trunk dendrites at  $0.93 \pm 0.13\ \mu\text{m}$  and targeting dendrites at  $0.64 \pm 0.01\ \mu\text{m}$ . The trunks of interneurons, therefore, appeared to be thicker than that observed by Morgan & Lichtman (2020) in our preparation.

To evaluate whether the two, trunk and targeting, classifications of dendritic morphology suited the diameter measures, we used Gaussian mixture modelling (GMM) to perform unsupervised fitting on all unclassified diameter measures (Reynolds, 2009). First, for a specified  $k=2$  number of probability densities, resultant cluster means were detected at 1.40

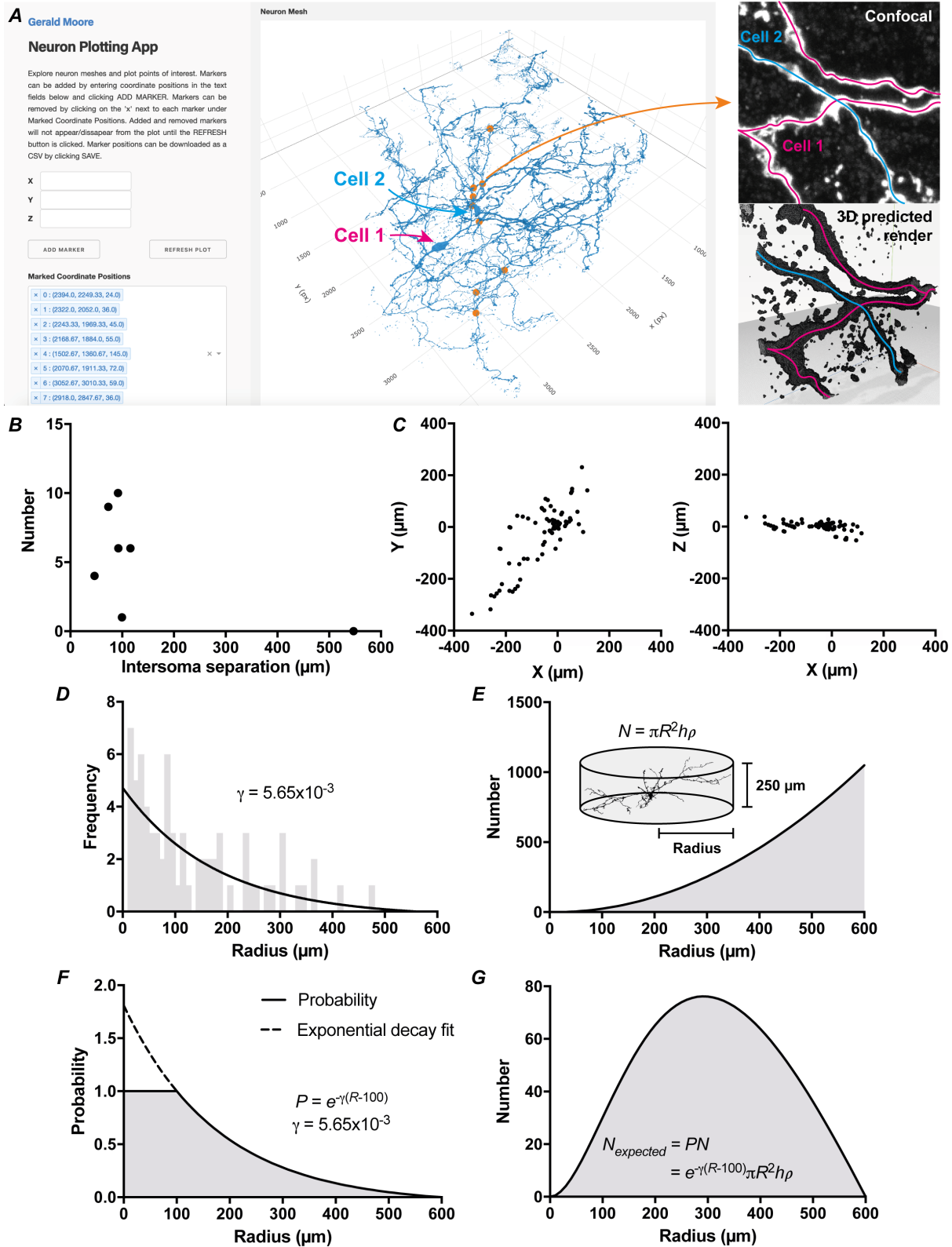
$\pm 0.77 \mu\text{m}$  (mean  $\pm$  SD,  $n=770$  measures across 6 interneuron pairs) for thicker dendrites and  $0.52 \pm 0.17 \mu\text{m}$  for the thinner dendrites. These measurements were in agreement with mean measures for the trunk and targeting dendrites. When naïve to the number of clusters, GMM returned an optimum  $k=4$  probability densities for  $0.49 \pm 0.14 \mu\text{m}$  (mean  $\pm$  SD),  $0.98 \pm 0.22 \mu\text{m}$ ,  $1.73 \pm 0.35 \mu\text{m}$  and  $3.82 \pm 0.83 \mu\text{m}$ , suggesting more morphological classifications of the dendrites. However, the thicker  $3.82 \pm 0.83 \mu\text{m}$  dendrites could be more representative of the direct arborisation from the soma or representative of the varicosities, discussed below. The other three cluster means could represent the trunk dendrites, medium thickness and more directed dendrites, and thinner more tortuous targeting dendrites.

There were many arborisations from the trunk that remained thick and long with little circuitousness, whilst some targeting dendrites showed extreme circuitousness with almost toroidal like wrapping. Dendrites that exhibited high tortuosity also had multiple varicosities as large as  $1\text{-}2 \mu\text{m}$  separated along the dendrites by a distance of  $3.84 \pm 2.20 \mu\text{m}$  (235 measures in the XY projection). Distances between swellings were measured in the XY projection, where they were more easily discerned, and where swellings appeared to be drastically larger compared to the surrounding neurite morphology. The thicker trunk dendrites also showed varicosities, but much reduced in frequency, and were harder to identify as they were already very thick even without the swellings. These varicosities could coincide with triadic terminal sites which are known to create large complexes.





**Figure 5.7.** Segmenting paired interneuron biocytin fills using deep learning. Deep learning (blue) compared to Otsu's thresholding (magenta). Individual interneurons show complex morphology with numerous circuitous neurites which spanned the entire 250  $\mu\text{m}$  depth of the acute slice preparation, plus spanning around 500  $\mu\text{m}$  in the coronal plane. Using a trained U-Net on manually segmented neurites, the complex neurite structure could be more accurately segmented compared to traditional thresholding methods. In particular, finer sub-micron neurites were preserved in segmentation (i-iii). Arrow markers indicate corresponding points in original (black), prediction through deep-learning (blue) and Otsu's thresholding (magenta). Scale bar 100  $\mu\text{m}$  in overview panels, and 10  $\mu\text{m}$  in enlarged panels.



**Figure 5.8.** Detecting putative interneuron-interneuron anatomical connectivity. **A.** Using the predicted segmentation of the neurite structures, interneurons are converted into a mesh using a cube marching algorithm (Lewiner et al., 2003), and decimating down to a smaller file size using the Blender 3D graphics software. Meshes are plotted using custom-built Plotly and

*Dash framework web application, allowing appositions of  $<5\ \mu\text{m}$  separation between paired interneurons to be visually landmarked (left). Marked positions were used to inspect the original high-resolution data set for validation (right). **B.** Several putative connections between interneuron pairs at indicated inter-soma distance, marked with evidence of connectivity from electrophysiological recordings. **C.** Locations of anatomical connectivity with respect to some centre (0, 0) projected in the XY plane (left) and ZX plane (right). **D.** Frequency of putative connections given distance from soma parameterised by exponential decay model with  $\gamma = 5.65 \times 10^{-3}$ . **E.** Projected density of nearby interneurons within a cylindrical volume of height  $250\ \mu\text{m}$  (equivalent to the slice thickness), given known density  $3.84 \times 10^{-6}\ \text{cells}/\mu\text{m}^3$  in the LGd. Number of interneurons,  $N$ ; radius,  $R$ ; height,  $h$ ; density,  $\rho$ . **F.** Parameterised probability of an interneuron observing connection with a neighbouring interneuron, using known putative connectivity within  $100\ \mu\text{m}$  ( $n=6$  interneuron pairs), and dendritic spanning limit at  $600\ \mu\text{m}$ . Probability,  $P$ ; radius,  $R$ . **G.** Combining predicted connectivity between neighbouring interneurons and modelled density of interneurons to provide an approximate model of the number of potential interneurons which may be connected to a single interneuron.*

### **Paired interneuron anatomical connectivity**

Paired interneurons from patch-clamp recordings were separated by an average inter-soma distance of  $152 \pm 175\ \mu\text{m}$  (mean  $\pm$  SD,  $n=6$  interneuron pairs) with one erroneously filled interneuron separated with a distance of  $547\ \mu\text{m}$  from a pair recorded in the same slice preparation. Excluding this from the average inter-soma distance gave a mean separation of  $86 \pm 23\ \mu\text{m}$ . Close interneuron separation could be attributed to the requirement to patch nearby cells during a paired recording protocol. All close pairs ( $n=6$ ) showed evidence of close appositions where respective neurites approached  $<5\ \mu\text{m}$  from a neurite of the other interneuron. The erroneously filled interneuron separated by  $547\ \mu\text{m}$  from the nearest of the two filled interneurons showed zero close appositions with the other two interneurons in the slice.

Close ( $<5\ \mu\text{m}$ ) appositions between interneuron pairs indicated preferential connectivity  $<100\ \mu\text{m}$  radius from the soma (39 of 72 measures; 54%) over connections in the distal regions of the dendritic tree  $>100\ \mu\text{m}$  from the soma (33 of 72 measures; 46%).



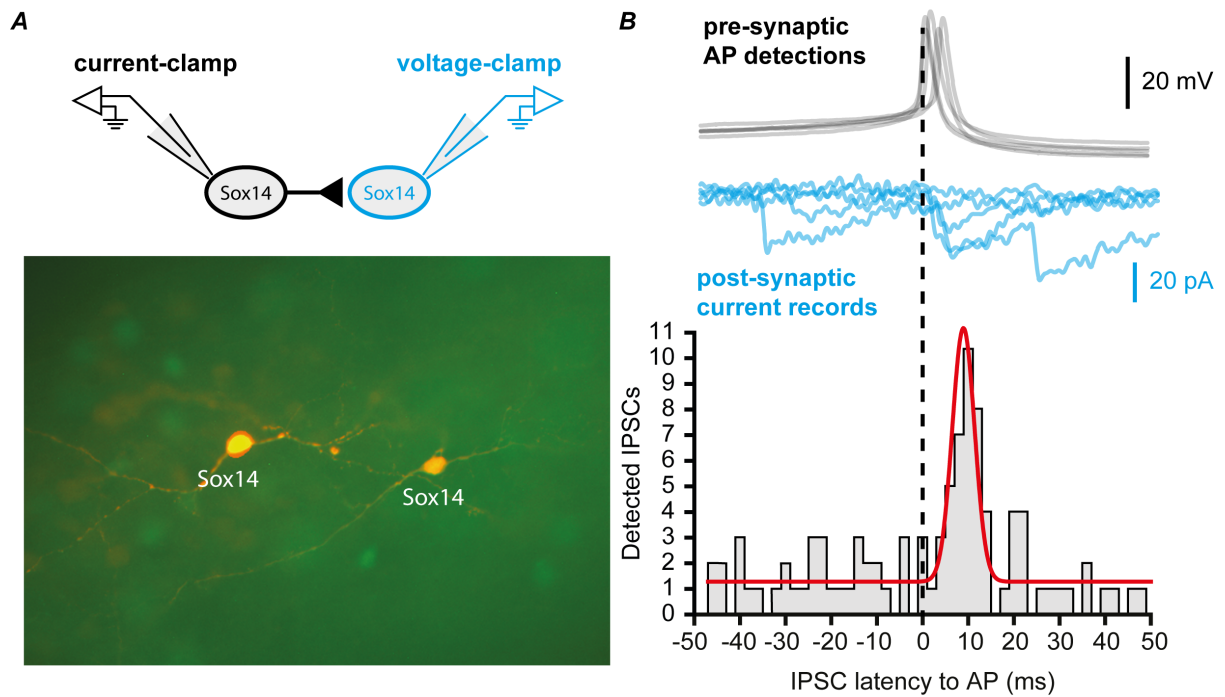
Across all paired interneurons, appositions occurring within 100  $\mu\text{m}$  Euclidean radius of the soma were more frequent (39 of 72 measures; 54%), than connections which were established in the more distal regions of the dendritic tree (33 of 72 measures; 46%). Appositions appeared to be distributed in a bipolar distribution (Fig. 5.8C) possibly indicative of the bipolar morphology of the interneurons. In the rostral-caudal plane, putative connections between interneurons were much more confined than in the medial-lateral plane with all points occurring within a Z-range of  $\pm 53.2 \mu\text{m}$  from the soma.

Parameterising an exponential decay to the frequency of close apposition with radial distance gave a decay constant  $\gamma = 5.65 \times 10^{-3}$ , with zero close appositions detected greater than 600  $\mu\text{m}$  from the soma (Fig. 5.8D). As the interneurons spanned a 500-600  $\mu\text{m}$  diameter, no appositions were expected to be detected outside of this area. The single interneuron with a 547  $\mu\text{m}$  inter-soma separation was used to impose a connectivity limit. Given a known interneuron density of  $3.84 \times 10^{-6}$  cells/ $\mu\text{m}^3$  determined from automated cell detection in the LGd of the Sox14gfp/+ mouse (n=3 mice), the number of interneurons within a radius of a target interneuron could be approximately modelled with a cylindrical volume of height 250  $\mu\text{m}$  (Fig. 5.8E). With a known 100% likelihood that neurons within a separation distance of <100  $\mu\text{m}$  observe close appositions (evidenced in all 6 interneuron pairs), and a known exponential parameterisation of the likelihood of an interneuron to observe appositions with a neighbouring interneuron, the probability of an interneuron observing appositions at a given radius can be parameterised by the same exponential decay with  $\gamma = 5.65 \times 10^{-3}$ , and capped with a 100% probability when within a 100  $\mu\text{m}$  radius (Fig. 5.8F). Combining this with the modelled density of neurons (Fig. 5.8E), the predicted number of appositions with neighbouring interneurons can be modelled by multiplying the neuron density by the probability of observing appositions (Fig. 5.8G). If using close appositions as a putative representation of anatomical connectivity, there as a very high number of modelled possible connections between interneurons in the LGd, with as many as 76 potentially connected interneurons at a radius of 300  $\mu\text{m}$ . However, the accuracy of this estimation is limited by several factors. First, interneuron dendrites do not fill a cylindrical volume in the slice, but instead, follow quite circuitous paths which were not guaranteed to take them close to neighbouring interneurons. The model also used putative connectivity between paired

interneurons as a proxy for connectivity between other interneurons. From the imaging data, there was no way to confirm whether these close appositions were synaptic terminals.

### **Paired interneuron functional connectivity**

To assess functional connectivity between interneurons as suggested by the evidence of putative anatomical connections, whole-cell current-clamp recordings were made on presynaptic Sox14 cells in the LGd of the Sox14gfp/+ mouse, with simultaneous voltage-clamp recordings on the neighbouring Sox14 cells (Fig. 5.9A). Evoked IPSCs in the postsynaptic Sox14 cell followed AP detections in the pre-synaptic interneuron after 10 ms. Although evidence of functional connectivity was present in the majority of paired recordings, the strength of connectivity was variable with 8-of-11 recorded pairs showing some degree of linear correlation (with a Pearson's  $R > 0$ ) between detected IPSCs in the postsynaptic interneuron and the pre-synaptic AP firing rate (Fig. 5.9B). 3-of-11 recorded pairs showed no or negative correlation between pre-synaptic APs and evoked IPSCs. This suggests that although anatomical and functional connectivity exists, the level of connectivity between interneurons may range between strong and weak levels of connection. Finally, when considering the 6 pairs of Sox14 neurons that were morphologically reconstructed, only 2 of the paired recordings associated with these cells showed evidence of strong functional connectivity as evidenced by the analysis illustrated in Fig. 5.9.



**Figure 5.9.** Paired interneuron recording in the LGd. **A.** Whole-cell current-clamp recording on presynaptic Sox14 interneuron in the LGd and whole-cell voltage-clamp recording on neighbouring postsynaptic Sox14 interneuron (top). Interneurons were filled with biocytin for visualisation and anatomical measures used in morphological analysis (bottom). **B.** Pre-synaptic APs (in black, top) and evoked postsynaptic IPSCs (in blue, top). Peristimulus time histogram (bottom) shows pre-synaptic AP evoked IPSCs in the postsynaptic interneuron, establishing paired interneuron functional connectivity in the LGd.

## 5.4 DISCUSSION

Interneuron number and distribution across the thalamus varies between species and possibly with respect to the complexity of the excitatory thalamus (Arcelli et al., 1997). Smaller brained mammals show restriction of the thalamic interneuron population to the visual thalamus, whilst higher mammals, such as primates, show a widespread interneuron distribution across the thalamus. Here it was suggested that interneuron presence could be far more widely distributed across the mouse thalamus than previously appreciated, with many first-order relay and higher-order nuclei showing the presence of Sox14+ cells. In a recent study, immunodetection of Sox14 and Gad showed 100% of the Sox14+ cells across the thalamic nuclei also co-expressed Gad, with first-order sensory nuclei such as LGd showing 98% representation of GABAergic interneurons with Sox14+Gad+ expression (Jager et al., 2020). Whilst there was also evidence of Sox14-Gad+ across the thalamus ( $22.1 \pm 4.0\%$ ), those cells appeared to cluster in spatially distinct regions, skewing towards higher-order nuclei. The abundant distribution of Sox14+ cells across the thalamus (Fig. 5.1, 3) and knowledge of the GABAergic profile of Sox14+ cells, suggests a widespread but low distribution of local thalamic interneurons, never previously realised. Investigating the function of these interneurons could reveal a more sophisticated computational paradigm for the rodent thalamus than previously thought.

Cortical layer 6 neurons in the primary visual cortex provided abundant corticogeniculate connectivity to local LGd interneurons as expected (Jurgens et al., 2012). These connections provide strong excitation to the interneurons, which in turn are recruited to inhibit TC relay cells. Corticogeniculate projections also appear to send collaterals to the RT. As shown here, GABAergic RT neurons establish inhibitory drive onto local LGd interneurons as well as relay cells. The functional significance of this has yet to be fully explored. The RT appeared to be highly involved in retinogeniculate signal processing, with evidence of recruitment by collaterals of ascending excitatory TC relay cell axons, and excitatory recruitment by collaterals of descending corticogeniculate axons. We then showed a functional level of connectivity between the RT and local LGd interneurons through optogenetic stimulation of the RT during whole-cell recording on Sox14 cells. The RT, therefore, appears crucial for the visual thalamus system and could play a role in synchronising signal propagation and attention modulation of retinogeniculate signals.

There was also evidence of labelling in the SC suggesting tectogeniculate projections to interneurons of the geniculate nucleus. However, the inability to distinguish between axons and ineffective targeting of rabies tracing to the LGd means these projections might not directly target LGd interneurons, but possibly those in the LGv or IGL. However, SC driver-like input to TC relay cells of the LGd has already been established (Bickford et al., 2015). Considering the complexity of monadic, dyadic and triadic synaptic motifs with interneurons, SC axons might also target interneurons.

Local interneurons showed evidence of putative anatomical connectivity between themselves which has only recently been anatomically shown in the mouse LGd (Morgan & Lichtman, 2020). Preference for putative interneuron to interneuron synapse sites was skewed towards the soma, with half of the putative connections occurring within 100  $\mu\text{m}$  of the soma. There also appeared to be a preference for connections within a  $\pm 53.2 \mu\text{m}$  range in the rostral-caudal plane. This may suggest connections in the rostral-caudal extent of the LGd are more restricted with connections in the medial-lateral and dorsal-ventral axis appearing more distally in the dendritic arborisation. Modelling potential connectivity between interneurons suggests each interneuron could be vastly connected to many others within a 500  $\mu\text{m}$  dendritic span. An exceedingly complex dendritic morphology with distal dendritic spanning across multiple highly functionalised domains within the LGd would also suggest a highly connected network. Recording pre-synaptic APs and postsynaptic IPSCs in paired interneurons showed that the anatomical connectivity was functionally present with presynaptic APs evoking postsynaptic IPSCs observed in 2 of 6 paired interneurons. Future studies are required to reveal the function of interneuron to interneuron connectivity, such a network might be required for synchronisation of temporal events through disinhibition (Pasik et al., 1976). Biocytin fills of interneurons in the LGd did not appear to spread to neighbouring interneurons suggesting lack of gap-junction coupling between interneurons (Supplementary Fig. 5.1).

Individual interneuron morphology was exceptionally complex, spanning a diameter of around 500  $\mu\text{m}$  in the coronal plane and projecting the complete 200-250  $\mu\text{m}$  anterior-posterior extent of slices during patch-clamp experiments. Dendrites could be visually

clustered into thicker trunk dendrites with diameter  $1.49 \pm 0.74 \mu\text{m}$  (mean  $\pm$  SD,  $n=189$  measures), and targeting dendrites with diameter  $0.52 \pm 0.19 \mu\text{m}$  ( $n=581$  measures) and agreed with a GMM of  $k=2$  clusters. However, there was evidence of three classifications of dendrite morphology with  $0.49 \pm 0.14 \mu\text{m}$  (mean  $\pm$  SD),  $0.98 \pm 0.22 \mu\text{m}$ ,  $1.73 \pm 0.35 \mu\text{m}$  and  $3.82 \pm 0.83 \mu\text{m}$ , although more measures are needed to confirm this. Targeting dendrites showed far more circuitous paths compared to the more direct trunks. However, tertiary branches of the dendrites did not all show circuitous neurite structure with some branches showing thin but still direct morphology.

In some cases, the circuitous morphology was extreme to the point of exhibiting toroidal like structures, which could be interneuron dendrites wrapping around dendrites of relay cells. Dendrites also showed high tortuosity with multiple varicosities suggesting sites of terminal contacts with RGC, relay cell, interneuron, RT projection or other input/output non-local neuron. The large spanning dendritic structure of interneurons makes them well suited to providing inhibitory input to many functionally distinct domains of the LGd. In addition, the type of dendritic morphology could be crucial for establishing specific connections with thicker and more direct dendrites relying on en-passant terminal formation whilst the targeting distal dendrites actively seek synaptic contacts with other cells. From a functional perspective, differences in dendritic morphology could also lend to different strengths of synaptic connectivity. Although paired interneuron connectivity was observed, there were varying degrees of strength.

This study demonstrates that interneurons have an immensely complex involvement into retinogeniculate, thalamocortical and corticothalamic visual information processing. Far from being simple inhibitory mechanisms to the local relay cells, interneurons are recruited by ascending and descending pathways either with direct terminal contacts or indirectly through recruitment of the RT, which then contacts the interneuron. Whilst the study by Morgan & Lichtman (2020) was comprehensive for a single interneuron, the research shown here establishes interneuron-interneuron connectivity to be functionally present with further evidence that LGd-wide dendritic spanning is characteristic of all interneurons. Connectivity which was determined to be from unknown sources by Morgan & Lichtman (2020), have also been shown to, in part, come from projecting GABAergic RT neurons which provides a level

of inhibition to local visual thalamic interneurons. The functional significance of this is for future investigation.

## 5.5 FUTURE PERSPECTIVES

One of the biggest challenges with mapping the interneuron input to the LGd was explicitly targeting the LGd without hitting the IGL and LGv. Furthermore, neurite GFP labelling was dense, making it difficult to distinguish individual axons once they hit the LG. This made it unclear as to which inputs were targeting local LGd interneurons. Sparse and targeted rabies tracing would, therefore, be more preferential for this investigation. This could be achieved by reducing the volume of primer and rabies virus. The injection's location could also be targeted to the most dorsal and anterior region of the LGd (i.e. the region of the LGd furthest from the IGL and LGv).

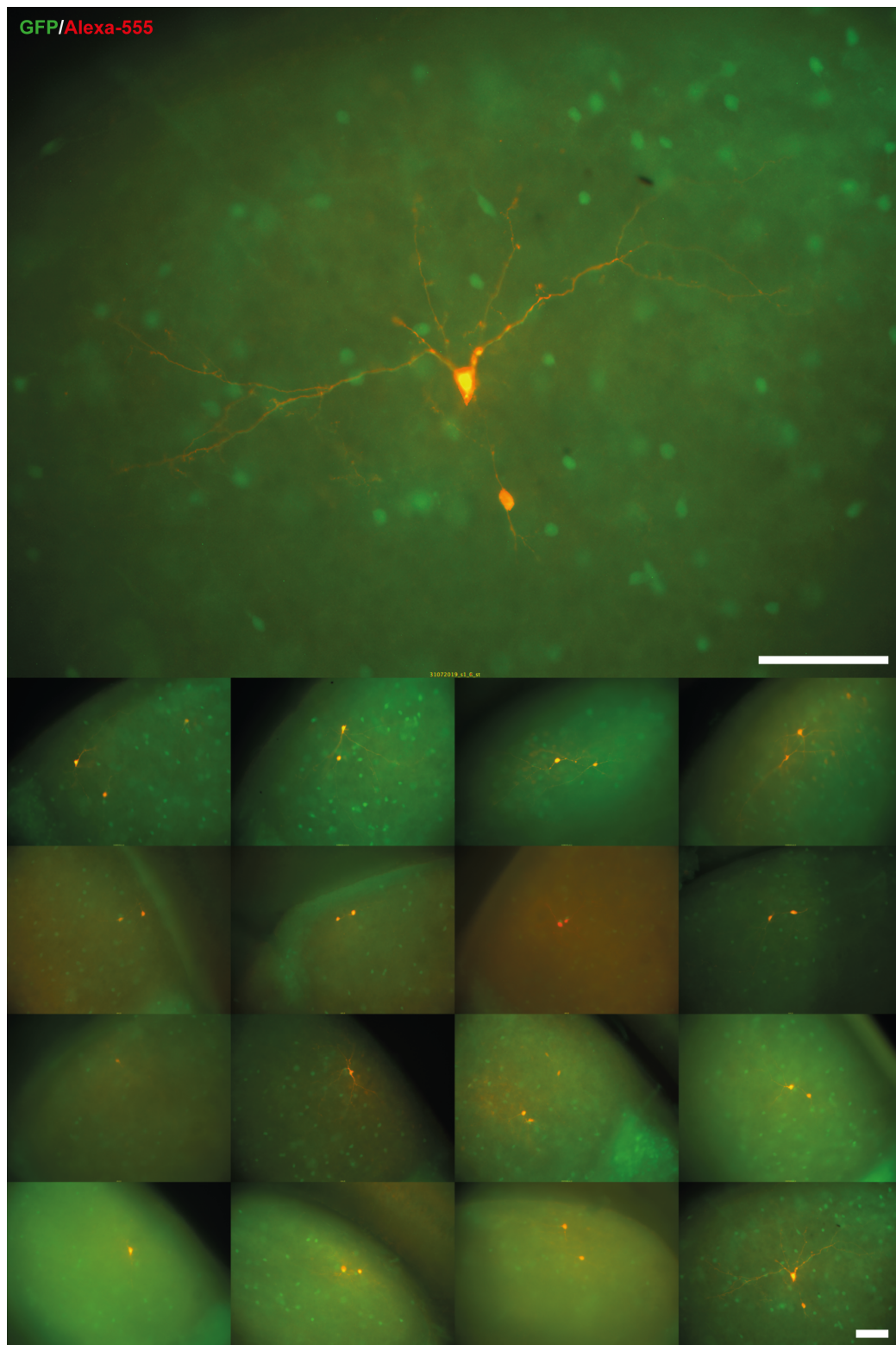
The interneuron connectivity model (Fig. 5.8B-G) is limited in several ways. First, the number of paired interneuron recordings was sparse, and therefore the inter-soma distance is limited to an average  $86 \pm 23 \mu\text{m}$ , excluding the erroneous fill. More paired interneuron recordings are therefore required, and explicitly more varied inter-soma separation is needed to truly establish the anatomical connectivity between interneurons with respect to their separation. Second, the proposed level of connectivity is modelled from a known density with a spherically spanning dendritic structure. This is not an accurate model of dendritic structure, as interneurons show a bipolar morphology. A Monte Carlo simulation could be better suited to modelling a more accurate level of dendritic spanning and therefore connectivity. This approach was used to show that the circuitous targeting dendrites appear to be preferentially designed to establish connectivity more than random chance would allow (Morgan & Lichtman, 2020).

The deep-learning approach for neurite segmentation was valuable for compressing the complex morphology of paired interneurons into a standard which can be more easily visualised. However, it was still challenging to distinguish between interneurons, given the complex intertwining of neurites. Single fills or dual dye labelling of interneurons would allow more accurate differentiation of the interneurons. Furthermore, this would permit manual tracing and more accurate measurements of the trunk, targeting dendrites and tortuosity. The mesh generation using the cube marching algorithm could also be improved by ignoring vertices and face placement under specific conditions to limit noise.



Of course, deep-learning for automated neuronal segmentation is very exciting. More work could go towards improving the model, including increasing training data and exploring ways to break the segmentation down into primary, secondary and tertiary neurites, and how rules on arborisations could help decipher between multiple labelled neurons in the same data set.

## 5.6 SUPPLEMENTARY DATA



**Supplementary Figure 5.1.** Widefield imaging of Alexa 555 conjugated biocytin fills of interneurons. Paired recordings from interneurons of the LGd (red) projected over GFP signal from all Sox14+ interneurons(green). Scale bar 100  $\mu\text{m}$ .

**AGE-RELATED CHANGES IN COGNITIVE PERFORMANCE AND BRAIN ANATOMY IN C57BL/6J  
MICE FROM YOUNG TO OLD AGE**

**ABSTRACT**

Healthy ageing leads to cognitive decline, structural changes in the brain and greater risk of neurodegenerative disease. In many cases, age-related changes to the body can result in complete loss of independence. However, not everyone shows the same trajectory of cognitive decline with studies suggesting those who exhibit high cognitive ability at younger ages might, to some degree, maintain high cognitive function into older age. Here we track cognitive performance in C57BL/6J mice (female, n=58) across 4, 8, 12 and 18 month age points, using a combination of a novel nine-arm radial maze task, and object recognition and location tasks. Mice were assessed for performance on the radial maze task over 16-days per age point, using decision-making time, arm visitation errors, unique sequence length score, exploration ratio, cumulative arm time and velocity. We additionally used novel object recognition and location tests evaluated with 1- and 15-minute retention intervals at each age point. As a population, mice showed little-to-no significant changes in cognitive performance from 4 to 18 months of age across all measures, with decision-making time suggesting the only significant improvement with age (Kruskal-Wallis non-parametric testing,  $p=0.0153$ ,  $n=58$ ). Although the significance did not hold true when evaluated for multiple comparisons between each age point (post hoc Dunnett's testing). We tracked individual performance using standardised Z-scores for each measure, showing significant positive correlation and predictability between 18-month performance and 4-month performance for decision making time ( $R^2=0.3495$ ,  $p=0.0038$ ,  $n=22$  mice), errors ( $R^2=0.3909$ ,  $p=0.0019$ ,  $n=22$ ), and unique sequence length score ( $R^2=0.6166$ ,  $p<0.0001$ ,  $n=22$ ). Combining Z-scores for unique sequence length, cumulative arm time and velocity revealed a small improvement to predictive power ( $R^2=0.6239$ ,  $p<0.0001$ ,  $n=22$ ). Lastly, MRI segmentation and high-resolution serial two-photon imaging of the prefrontal cortex (PFC) were used to illustrate how anatomical changes in the brain can be tracked and correlated with behavioural measurements.

## 6.1 INTRODUCTION

Structural and functional changes in the brain present two key facets of healthy ageing, which can ultimately impact behaviour and cognitive performance. Short- and long-term ageing studies have revealed a decline in episodic, semantic and working memory performance (Deary et al., 2009; Salthouse, 2012). Affects to general mental health have also been observed with an increase in depressive disorders, of which there appears to be an increased prevalence in women (Lukaschek et al., 2017). Cognitive ageing, meaning performance decline in several cognitive domains, presents perhaps the most important and investigated changes with age, most likely on account of utility in everyday life. Lamentably, in many cases, cognitive decline can be severe to the point of partial or complete loss of independence. The population of those aged over 65 are more vulnerable to health-related issues, stemming from the normal ageing process and increased prevalence of the disease. In the UK, 7% of those aged over 65 suffer from dementia, and the risk of acquiring dementia increases over 80 years of age from 7% to 17% (Prince et al., 2014). Home-based care of those aged over 65 years is projected to significantly increase by over 80% by 2035 in England (The King's Fund, UK). With a projected 16% of the world's population aged over 65 by 2050 (United Nations, 2019), it is becoming increasingly important to understand how we may track, predict and identify those at risk of severe age-related deterioration. Furthered understanding can help treatment design and intervention if needed.

There are two facets to cognitive intelligence; fluid intelligence which concerns reasoning and novel problem solving, and crystallised intelligence which concerns the accrument of knowledge and skills (Cattell, 1987). Throughout ageing, performance in both these facets appears to change. A study of containing over 3000 adults between 30 and 70 years of age indicates that fluid intelligence reduced by -0.02 standard deviations per year. On the other hand, crystallised intelligence showed the opposite trend with an increase in 0.02 standard deviations in performance per year (Salthouse, 2012). This is supported by a cross-sectional study of those in the 35 to 80 year age range showing a consistent decline in memory performance (Rönnlund et al., 2005).

However, ageing studies can often be disagreeable, with variability in lifestyle and cross-sectional or longitudinal study design yielding different results (Salthouse, 2014). Education

and societal changes occur over multiple decades, meaning a 50-year age gap study between young and older participants could be more reflective of uncontrollable environmental factors rather than meaningfully related to a healthy ageing process. For example, a longitudinal study on episodic memory over the 35 to 80 year concluded that episodic memory performance remained unchanged until as late as 60 years of age, before appearing to decrease, conflicting with the earlier statement (Murman, 2015; Nilsson, 2003). Emotional state which can be related to external environmental factors could also play a role in the severity of episodic memory performance loss (Kinugawa et al., 2013). Cross-sectional studies typically reveal monotonic decrease in cognitive measures, whilst longitudinal studies can reveal stable or incremental performance (Salthouse, 2014). Interestingly, cognitive studies which use subjects from the same birth cohort but only test once at different years show more agreement with cross-sectional studies (Rönnlund et al., 2005). Repeated measure studies at successive age points could impact cognition measurements and explain why age-related trajectories appear to differ between different study designs (Salthouse, 2014).

The idea that higher intelligence offers a level of protection against cognitive decline has been explored since the turn of the millennia (Stern, 2006). Level of intelligence, education and occupational attainment has been reported to influence the risk of dementia with subjects possessing low educational and occupational attainment suffering highest dementia rates in older age, compared to those who possessed only one of those elements (Lenehan et al., 2015; Stern et al., 1994; Whalley et al., 2000). Occupation also appears to present one of the strongest associations with later age cognitive decline (Singh-Manoux et al., 2011; Stern et al., 1994). Additionally, young adult language ability seemed predictive of later life cognition and risk of Alzheimer's disease (AD) with high early life linguistic skills leading to a reduced cognitive deficit and delayed AD onset (Snowdon et al., 2000). Interestingly, individuals presenting with AD and dementia appear to suffer increased cognitive decline when demonstrating high cognition (Andel et al., 2006). High early-life attainment could, therefore, delay onset of the disease but result in rapid cognitive decline once onset occurs.

Extensive longitudinal studies in humans are generally sparse. However, some studies performed from age 11 up to mid-60s collectively appear to show a correlation between an individual's performance at a young age versus their performance at old age (Mccall, 1977;

Owens, 1966; Schwartzman et al., 1987). In addition, Lothian cohort studies set up from the 1920s to 1950s were designed to track how cognition and physical and mental health changes in subjects from childhood to old age (Deary et al., 2000, 2004, 2009, 2013). Such studies present the most comprehensive assessment of an individual's trajectory throughout ageing, with strong evidence that performance in the Moray House Test at 11 years of age is correlated with their performance at 90 years of age (Deary et al., 2013). The Moray House Test was specially designed to test multiple mental attributes, with questions covering verbal reasoning, numerical arithmetic, word classification, following directions, cypher decoding and many others. This early life assessment demonstrates validity for potentially predicting how an individual may perform later in life.

Alongside behavioural cognitive changes, we also observe physical anatomical alterations in the brain. Magnetic resonance imaging (MRI) studies have shown grey matter volume loss from as early as 20 years of age when studied in those aged between 20 and 86 years of age (Ge et al., 2002). In addition, atrophy appears to be specific to brain regions of which the prefrontal cortex (essential for complex cognitive function) and cerebellum (important for motor function) show the highest volume loss (Terribilli et al., 2011). Interestingly, there are reports that temporal regions of the brain show an increase in volume,==== although one study was inconclusive in stating whether this was a true increase or a non-significant change (Terribilli et al., 2011). Notably, the presence of neurodegenerative disease also leads to brain volume atrophy with Alzheimer's leading to volume reduction in temporal regions of the brain, possibly suggesting Alzheimer's could lead to volume loss in high order association regions in contrast to a normal ageing atrophy process (Frisoni et al., 2002). This rapid change in structural anatomy could correlate with an observed rapid decline in cognition following the onset of disease.

White matter also suffers from atrophy with age and could potentially occur at a more rapid rate with a study on those between 30 and 90 years of age showing white matter degradation at twice the rate of grey matter loss (Jernigan et al., 2001). Up until 50 years of age, white matter loss could also accelerate (Raz & Rodrigue, 2006). In contrast, one study finds that white matter volume increases until a peak at 50 years at which point volume decreases (Liu et al., 2016). The same study also noted significantly higher white matter volume in females

over males. However, this result appears inconsistent compared to other sex-related aging studies which have shown that males have greater correlation of white matter volume with age (Allen et al., 2005), or that women show greater reduction in white matter volume (Voineskos et al., 2012).

Areas displaying white matter atrophy also do not appear to directly correlate with grey matter atrophy (Meier-Ruge et al., 1992). A notable pathology of white matter degradation are white matter hyperintensities (WMH) which are lesions usually attributed to degeneration of small vessels. Women tend to have an increased prevalence of WMH with higher occurrence in subcortical regions and increased severity of lesions (De Leeuw et al., 2001). Increased WMH burden also correlates with reduced cognitive performance, particularly with regards to executive function (Gunning-Dixon et al., 2009; Gunning-Dixon & Raz, 2000). As myelin insulation facilitates rapid electrical conductance, WMH also impacts reaction times and cognitive processing speeds (De Groot et al., 2000, 2002). That said, there are still many discrepancies between pathogenic factor, pathology and clinical manifestation, with incongruous relationships between the age, severity and volume of WMHs, and the extent of cognitive impact (Brickman et al., 2011). It is suspected the theory of cognitive reserve could help explain why there is so much observed variability between individuals.

Assuredly, individuals can show evidence of beneficial anatomical changes in the brain which can help maintain high levels of cognition. Functional MRI studies of older subjects during cognitive tasks show those who present poor cognitive function also present patterns of under-activation in the frontal cortex (Bartzokis et al., 2001; Emery et al., 2008; Ram et al., 2005). However, older subjects with higher levels of cognitive performance show hemispheric asymmetry reduction, termed the HAROLD mechanism (Cabeza et al., 2002; Höller-Wallscheid et al., 2017). Such a mechanism could be deployed to counteract the effects of age-related cognitive decline. Demonstrably, older subjects that performed comparatively to younger subjects in cognition tasks appeared to show bilateral recruitment of homologous brain regions. In contrast, the younger subjects showed asymmetric activation lateralised to only a single hemisphere. When suppressing activation in targeted brain regions through transcranial magnetic stimulation (TMS), suppressing the overactive regions in older subjects

appeared to lower cognitive ability (Rossi et al., 2005), whilst eliciting activity enabled older subjects to maintain higher cognitive performance (Sole-Padulles, 2005).

Despite the indisputable importance of ageing studies in humans, few studies seek to answer similar questions using animal models, such as mice, which offer more convenient study design and controllable environment and enrichment levels. Furthermore, a limited number of studies seek to validate animal models of ageing in equivalence to what we observe in humans. However, studies testing spatial memory using the Morris water maze and novel open-field test have shown small but non-significant changes in performance between mice when evaluated across 3 to 20 months (King & Arendash, 2002; M. Weber et al., 2015). Such insignificant observations could speak more to the sensitivity of the tests. The water maze, for example, is valid for testing allocentric spatial learning and working memory, but is high anxiety-inducing and less sensitive in long-term age-related studies where mice can effectively learn the task in early ages (Hodges, 1996). A novel object location study in C57BL/6J mice revealed that both male and female mice suffered in cognitive performance in later ages. However, this decline was not observed in novel object recognition tasks, but in the novel object location task (Benice et al., 2006). Female mice also appear to suffer more overall behavioural deficits, with observations of reduced passive avoidance behaviour (Benice et al., 2006). In contrast, the same study observed no differences between sex and task performance for mice which were young or middle-aged. Instead, only mice between 18 and 20 months showed inclination of cognitive decline. Crucially, many age-related studies could be confounded by their cross-sectional design. Widely used inbred strains such as C57BL/6J mice are commonly used for behavioural studies based on their presumed genetic homogeneity. However, several studies have shown that there are extensive genomic variations within inbred strains and sublines, even of the C57BL/6J line (Navarro et al., 2012; Simon et al., 2013).

Longitudinal studies in mice are surprisingly lacking and limited to simple measure tasks. To this end, we performed a long-term age-related study using a novel nine-arm radial maze design (Abuhamdah et al., 2016). The novel design expanded on the already established radial maze design with demonstrable use for long-term age-related studies (Hodges, 1996). The radial maze design also afforded the ability to measure multiple behavioural parameters as a



proxy for cognitive performance, including the decision making time between arm visitations, errors in exploring arms, number of unique arm visitations, rate of learning the task, exploration ratio of the maze, time spent on the arms and the velocity. Mice subjected to the maze were initially exposed to the task at 4 months over a total of 16 days, with repeated measurements at 8, 12 and 18 months of age, allowing the longitudinal impact of cognitive behaviour to be assessed.

In addition to the radial maze task, we performed novel object location and recognition tasks to assess spatial memory performance. Manual event logging and video tracking were used to track population changes and to investigate individual animal performance and early life predictivity of late-life performance. Finally, we evaluated the use of MRI and serial-two photon (STP) imaging technology for measuring changes in white matter and grey matter volume and changes in the connectivity level of intratelencephalic (IT) neurons projecting across the prefrontal cortex (PFC) of the brain (Dembrow & Johnston, 2014); an area which is heavily recruited during demanding cognitive tasks and for spatial memory (Goyal et al., 2008).

## 6.2 METHODS

The methods and work presented here represent the efforts of multiple researchers; Rushdie Abuhamdah (Imperial College London), Gerald Moore (Imperial College London), Deyl Djama (Imperial College London), Eugene Kim (King's College London), and Diana Cash (King's College London). Credit is acknowledged in the relevant methods section.

### Animals

For the radial maze and object location and recognition experiments, fifty-eight C57BL/6J (Charles River, UK) mice were housed 4 per cage under a 12-hour light:dark cycle (light 0700–1900 hours) at  $23 \pm 1$  °C with ad libitum access to food and water. Mice were left to acclimatise in their housing conditions for two weeks prior to experiments. Animal treatment and husbandry were carried out in accordance with the Animals (Scientific Procedures) Act 1986, UK.

For injections into the PFC using retrograde virus labelling, 12-month-old C57BL/6J (Charles River, UK) mice were used and housed in the same conditions as above.

### Nine-arm radial maze

The radial maze was constructed from 5 mm thick grey PVC and consisted of a central platform with nine bridges ( $15.2 \times 11.2$  cm) at 40° incline, radiating from the edges of the central platform (Fig. 6.1A). Nine arms ( $35.0 \times 11.2$  cm) were fixed to the ends of each respective bridge with a unique visual cue positioned on the end of each arm ( $20.2 \times 11.2$  cm). The cue could only be observed if the mouse had fully climbed the bridge. Each bridge featured a metal mesh floor to aid the mice on the incline, and a partial narrowing to the bridge entryway was made using a small, transparent ( $9 \times 6$  cm) barrier. This was done to prevent mice from making inattentive arm visitation choices by merely traversing the next adjacent arm. At the end of each arm and beneath the visual cue, a small food pellet was placed to incentivise food-deprived mice (see below) to explore the arm. The maze was placed in the centre of an animal behaviour room with the central platform raised 70 cm above the ground and was surrounded by a white curtain to prevent mice recognising external cues in the environment.

Testing with the radial maze (RM) was split into four stages (I-IV), respectively referred to as RMI, RMII, RMIII and RMIV, with cohorts of mice and testing ages arranged as follows. RMI consisted only of 4-month mice (n=16). RMII testing occurred 4 months after RMI and consisted of 8-month mice (n=16) and 4-month mice (n=16). RMIII testing occurred 4 months after RMII and consisted of 12-month mice (n=16), 8-month mice (n=16) and 4-month mice (n=16). Lastly, RMIV testing occurred 4 months after RMIII and consisted of 18-month mice (n=16), 12-month mice (n=16), 8-month mice (n=16) and 4-month mice (n=10). Due to varying degrees of longitudinal testing for some cohorts, some mice reached endpoint before completing all testing stages. Relevant numbers are reported in the corresponding 'Results' section.

At each stage, mice were subjected to 16 days of radial maze testing with the following exact testing ages; 4 months (post-natal age in days [P] P125-P140), 8 months (P257-P272), 12 months (P385-P400) and 18 months (P566-P581). One day prior to each RM testing stage, mice were individually weighed, and food-restricted. Weight monitoring was continued throughout the testing stage to ensure mice did not drop below 90% of their baseline, in accordance with animal welfare protocol regulations.

On each test day, mice were introduced into the central platform of the maze using a beaker and left to explore/complete the task. The task was concluded if mice satisfied either of the following two requirements; mice had explored a total of any nine arms with arm exploration classified by placing both front and hind feet onto an arm, or they spent a total of 10 minutes exploring the maze. To prevent mice from associating locations of the maze with task completion and ensuing removal from the maze, we allowed mice to return to the centre of the maze before removing them from the arena. Mice were returned to their cage, and the maze was cleaned with disinfectant to remove odour.

All mice were video recorded (800 × 600 pixels, 14.8 frames per second) and tracked using EthoVision software (Noldus Information Technology, USA). A manual event log was used for recording the sequence of arms visited by each mouse plus events where the food pellet was consumed (partially or entirely) and recording the number of errors as defined by the number of unvisited arms.

The radial maze design, construction and experimentation were performed by Rushdie Abuhamdah, Imperial College London.

### **Open-field and novel object location and recognition**

A square open-field arena (80 × 80 cm), also used for object location and recognition, was constructed from 5 mm thick grey PVC with an enclosing wall (50 cm) to prevent mice from recognising external visual cues. The arena was split into three areas; the central area consisted of a 16 × 16 cm square in the centre of the arena, the inner area consisted of a 16 cm wide region surrounding the central area, and the outer area consisted of a 16 cm wide region and took up the remaining space between the inner area and the wall of the arena (Fig. 6.2C).

Mice tested in the radial maze were also subjected to open-field, novel object recognition and novel object location tasks, prior to starting the 16-day radial maze session. The order of testing for mice was 3 days in the open-field, 4 days in object recognition and location, and 16 days in the radial maze amounting to 23 days of testing overall at each age point. Open-field testing was done to identify any mice which showed high levels of anxiety or significantly reduced exploratory behaviour. During open-field testing, mice were placed into the central area using a beaker and allowed to explore the empty arena for a total of 10 minutes on 3 consecutive days. Mice were removed and the arena was sterilised with disinfectant between each mouse.

For novel object location and recognition testing, we placed novel and familiar objects at any of the four vertices adjoining the inner and outer area. They were therefore located 16 cm from the arena wall. During testing, mice were carefully placed onto the central area using a beaker and given three minutes to explore two familiar objects placed at two of the four possible object locations. Mice were removed for a 1- or 15-minute retention interval. For object location testing, one of the familiar objects was moved to a new location, presenting a novel object. For object recognition testing, one of the familiar objects was replaced with a novel object. Mice were returned to the arena after the retention interval and left to explore for a further 3 minutes. Following task completion, mice were removed and returned to their cages, and the open-field arena was cleaned with disinfectant. On successive days, mice were

returned to the arena to complete all variants of the task such that each mouse was subject to 1- and 15-minute retention interval for objection recognition and object location tasks.

All mice were video recorded (800 × 600 pixels, 14.8 frames per second) and tracked using EthoVision software (Noldus Information Technology, USA).

All open-field, novel object recognition and location tasks were performed by Rushdie Abuhamdah, Imperial College London.

### **Radial maze and objection location and recognition performance analysis**

Performance measures of mice were extracted from either the manual event log, or from video tracking results, and additionally converted into a standardised Z-score to permit comparison between individual animals.

*Decision-making time.* Extracted from the tracking data. Mean decision-making time,  $\overline{DMT}$ , was calculated with the following.

$$\overline{DMT} = \frac{T_{trial}}{16(n + 1)} \quad (\text{Eq. 6.1})$$

$T_{trial}$  represents the time spent in the trial (either 10 minutes or less if nine arms were visited),  $n$  represents the number of arms visited incremented by +1 so a baseline of zero arm visits results in a real  $\overline{DMT}$  value, and 16 represents the number of trial days. The  $\overline{DMT}$  for each mouse was converted into a Z-score with respect to the sample by subtracting the mean and dividing by the sample standard deviation. To correlate high Z-score with a lower  $\overline{DMT}$  the time for each animal was multiplied by -1 to invert the distribution, before calculating the Z-score.

*Number of errors.* Extracted from the manual event log. Mean error,  $\bar{E}$ , was calculated with the following.

$$\bar{E} = \frac{\sum E}{16} \quad (\text{Eq. 6.2})$$

$E$  represents the errors on each trial day, and 16 represents the number of trial days. The error rate was converted into a Z-score by subtracting the sample mean and dividing by the

sample standard deviation. To correlate a high Z-score with a low error rate, the error rates were multiplied by -1 to invert the distribution, before calculating the Z-score.

*Unique sequence length.* Extracted from the manual event log. In general, better performance was indicated by proxy of exploration and more unique arm visits. However, mice might engage more with the task at any point during their 10-minute exploration of the maze. Rather than limit the measure to the first unique sequence, and to be sensitive to engagement at any time in the task, all possible unique sequences were extracted from the arm choice where each unique sequence cannot already belong to another unique sequence. In this manner, only the largest unique sequences were found where the sequence was not limited to the start of the trial. Sequence length and length frequency across the 16-day session were found for each mouse. To correlate longer unique sequences with higher task performance, a weighted Euclidean sum square product score, *USL*, was calculated using the following.

$$USL = \sqrt{\sum_{i=1}^9 (if_i)^2} \quad (\text{Eq. 6.3})$$

*f* represents the frequency of sequence length *i*. The Z-score was calculated by subtracting the sample mean and dividing by the standard deviation for each mouse. As the sequence length weighted the score, mice with few long-length unique sequences and fewer number of unique sequences overall, achieved a higher Z-score than mice with a higher frequency of low-length unique sequences.

*Learning rate.* Extracted from the manual event log. Linear regression was used to determine the rate of change of error for each mouse across the 16-day testing session. Learning rate was calculated by multiplying the error rate by -1, such that high errors early in the test session with low errors later in test session correspond with high learning ability. This meant that high learning mice reduced the errors they accumulated over the RM testing period. The Z-score was calculated by subtracting the sample mean and dividing by the standard deviation for each mouse. A high Z-score corresponded with higher learning rate.

*Exploration amount.* Extracted from video tracking. The radial maze was split into 28 regions; 9 arms, 9 bridges, 9 outer-centre and 1 inner-centre. Video tracking was used to measure how

many regions were explored by each mouse. The mean exploration amount was calculated as the sum of all explored regions divided by the total possible number. As mice were introduced to the maze using the inner-centre area during the trial, this region was omitted from the analysis, giving a total of 27 explorable regions per day. The Z-score was calculated by subtracting the sample mean and dividing by the standard deviation for each mouse. Mice with a higher exploration ratio of the radial maze had a higher Z-score.

*Cumulative arm time.* Extracted from video tracking. The mean total time each mouse spent on the arms. The Z-score was calculated by subtracting the sample mean and dividing by the standard deviation for each mouse. Mice with a higher cumulative arm time had a higher Z-score.

*Velocity.* Extracted from video tracking. Video tracking gave the total distance travelled by each mouse. This was divided by the time spent in the trial to get an average trial velocity which was then averaged across the 16 trial days. The Z-score was calculated by subtracting the sample mean and dividing by the standard deviation. Mice with higher velocities had a higher Z-score.

*OR1 preference ratio.* Following a 1-minute retention interval between the sample and choice phase, the novel object preference ratio was calculated with the following formula.

$$OR1 = \frac{T_{novel}}{T_{novel} + T_{familiar}} \quad (\text{Eq. 6.4})$$

$T_{novel}$  represents the time spent with the novel object during the choice phase, and  $T_{familiar}$  represents the time spent with the familiar object during the choice phase. This value was converted into a Z-score for each mouse by subtracting the sample mean and dividing by the standard deviation. Mice showing a preference for exploring the novel object obtained a higher Z-score.

*OR15 preference ratio.* Same as OR1 but using the object recognition 15-minute retention interval measures.

*OL1 preference ratio.* Same as OR1 but using the object location 1-minute retention interval measures.

*OL15 preference ratio.* Same as OR1 but using the object location 15-minute retention interval measures.

### **PFC retrograde labelling injections**

12-month old mice received stereotaxic injections of 0.3  $\mu$ l pAAV-hSyn-EGFP (retrograde) at 0.8  $\mu$ l/min flow rate into the infralimbic PFC at coordinates ML 0.33, AP 2.10 and DV -2.13. Mice were kept under anaesthesia with isoflurane for the duration of the procedure. Stereotaxic injection procedure was performed by Diana Lucaci, Imperial College London.

After 2 weeks, mice were anaesthetised with intraperitoneal (IP) injection of sodium pentobarbital (50 mg/kg) and transcardially perfused with 20 ml of 1X phosphate-buffered saline (PBS, pH 7.4) at 4°C, followed by 20 ml of 4% paraformaldehyde (PFA) at 4°C. After extraction, brain tissue was stored in 4% PFA at 4°C for 24 hours before washing three times with 1X PBS and storing in 1X PBS with 0.01% NaN<sub>3</sub> until imaging.

### **Serial two-photon imaging of retrograde labelled PFC neurons**

Prior to imaging, samples were embedded in a 4.5% oxidised-agarose solution as described in the 'Methods' of 'Chapter 2'. Mice were STP imaged using TissueCyte 1000 as described in 'Methods' of 'Chapter 2'. For imaging pAAV-hSyn-EGFP labelling, a 920 nm excitation wavelength was used in combination with GFP emission collected on channel 2 with a 525/40 nm Brightline bandpass filter (Laser 2000 Ltd.). Image tile data was stitched following the procedure detailed in 'Methods' of 'Chapter 2'.

### **Magnetic resonance imaging**

Selected mice that completed all radial maze, novel object location and novel object recognition tests were imaged with MRI (n=14) using a 7 Tesla pre-clinical MR system (Agilent Technologies, USA). Brains were first extracted following IP injection of sodium pentobarbital (50 mg/kg) and transcardial perfusion with 20 ml of 1X PBS (pH 7.4) at 4°C, followed by 20 ml of 4% PFA at 4°C. Brain tissue was stored in 4% PFA at 4°C for 24 hours before washing three



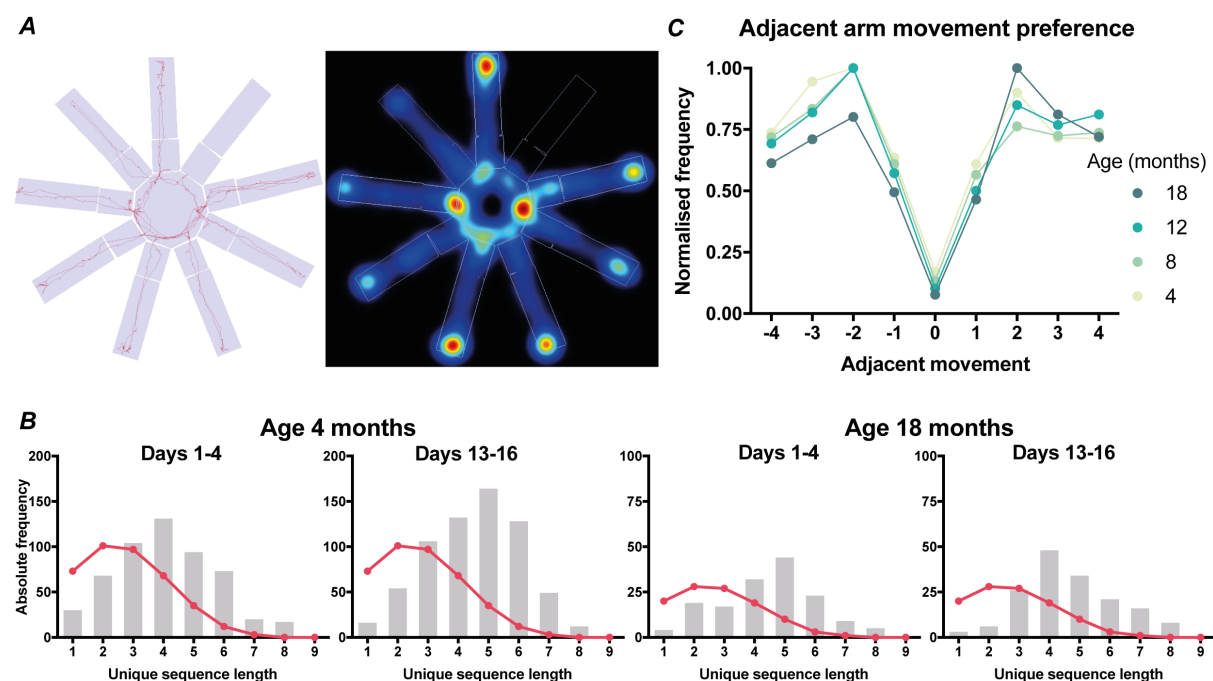
times with 1X PBS and storing in 1X PBS with 0.01% NaN<sub>3</sub> until imaging. Four brain samples were imaged at once by positioning them inside a falcon tube and separating them with a polypropylene divider. Samples were immersed in an inert fluorinated solution (Fomblin, Solvay) and imaged using a T1 (MP2RAGE), T2 (FSE) and diffusion (DTI) scan configuration with isotropic 1 mm<sup>3</sup> voxel size. MRI acquisition was made in collaboration with Diana Cash and Kim Eugene, King's College London.

### **MRI analysis**

MRI images were converted to NIFTI format and analysed using FMRIB Software Library (FSL; Jenkinson et al., 2012). For grey matter and white matter segmentation, FMRIB's Automated Segmentation Tool (FAST) was used on T1 data (Zhang et al., 2001). FAST has the benefit of being more robust, less sensitive to noise and also implements bias field correction to reduce low-frequency bias in the magnetic field. MRI segmentation was done by Deyl Djama, Imperial College London.

### 6.3 RESULTS

Due to the volume of video data across all experiments, not all video recorded animals were tracked with EthoVision. The number of animals used in the analysis for some measures is, therefore, lower compared to results retrieved from the manual event log. Respective numbers are reported where appropriate.



**Figure 6.1.** Nine-arm radial maze. **A.** EthoVision tracking of an exemplar mouse performing the task (left, the red line shows animal path) with a corresponding heat map (right). **B.** Frequency of any unique sequence length. Mice (grey bars) adopted a strategy that performed better than a stochastic model of the same number of mice (red line). Across all ages (only 4 months [ $n=58$ ] and 18 months [ $n=22$ ] shown here) mice accomplished a modal value of four or five unique arm choices before returning to an arm they had previously visited, incurring an error of one missed arm. A stochastic model achieved a modal unique arm sequence length of two. At 4 months, mice appeared to improve from the initial four days exposure to the task with modal unique arm sequence of four, to a unique arm sequence of five in the last four days of the task. **C.** Adjacent arm movement. Mice show an inclination to move to the second adjacent arm clockwise or anti-clockwise from the current arm position ( $91.44 \pm 0.04\%$  mean  $\pm$  SEM normalised across all ages). In contrast, mice were disinclined to visit the same arm ( $12.12 \pm 0.02\%$  normalised frequency across all ages), or the next adjacent arm ( $55.64 \pm 0.02\%$  normalised frequency across all ages).

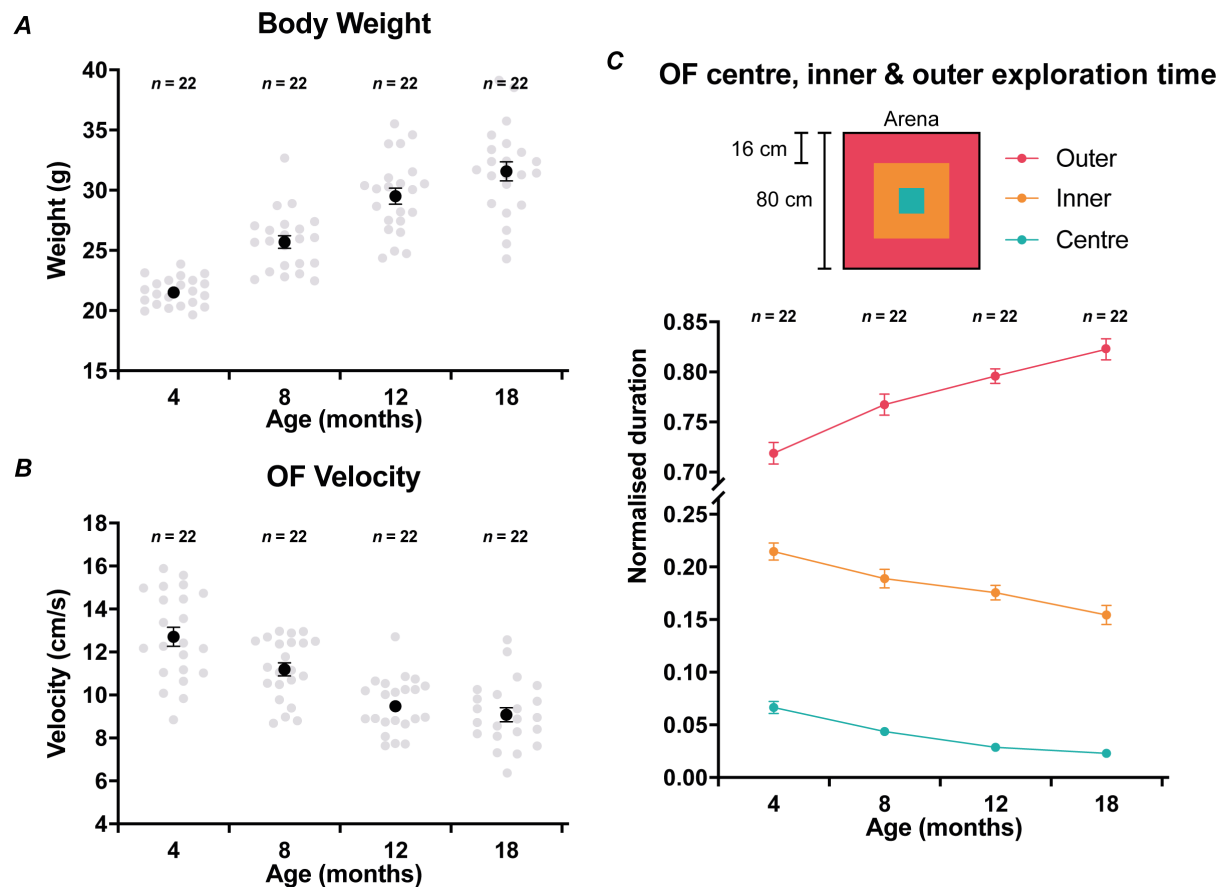
### **Mice adopted a strategy in the radial maze task with above random performance**

As a population, mice showed little anxiety with a tendency to explore the majority of the maze, as revealed by EthoVision tracking of an exemplar mouse (Fig. 6.1A). A heat map showing cumulative time spent by mice showed a tendency to spend time at the end of the arms and the bridge entrance (Fig. 6.1A). The centres of the arms were the most exposed region of the maze and therefore showed the least cumulative activity. In addition, mice preferred to explore the outer-centre area, as opposed to the more exposed inner-centre.

Mice appeared to adopt an exploration strategy which favoured unique exploration of the arms before making an error, as represented by a unique sequence length value (Fig. 6.1B). Here, unique sequence length covered all possible unique sequences in the arm choices and was not limited to the first unique sequence. Across all ages, mice achieved a modal unique sequence length between four and five arms. At 4 months when mice were first exposed to the maze, the modal value shifted from four unique arm choices in the first four days of testing to a modal value of five unique arms choices in the last four days of testing, suggesting an improvement in learned mice strategy (n=58). At 18 months, mice showed the opposite with a modal value of five unique arms choices in the first four days, reducing to a modal value of four at the end of the 16-day session (n=22). Additionally, mice performed better than a simple stochastic model of the same animal-number which randomly chose arm visitations with a modal value of two unique arm choices before incurring an error.

Strategy adoption can also be seen by looking at the preferred adjacent arm choice. For a nine-arm radial maze, there were nine possible choices to make from any current arm position. These corresponded with an adjacent movement of -4 to +4 with respect to clockwise or anti-clockwise movement, with the option of traversing the same arm again for an adjacent arm movement of 0. Mice preferred to move -2 or +2 adjacent arms with a mean of  $91.44 \pm 0.04\%$  (mean  $\pm$  SEM, n=16,347 total decisions from) frequency normalised to the maximum frequency across the respective age group. A normalised mean of  $55.64 \pm 0.02\%$  (mean  $\pm$  SEM, n=16,347 total decisions) of decisions lead mice to the next adjacent arm (-1 or +1) with a normalised mean of  $12.12 \pm 0.02\%$  (mean  $\pm$  SEM, n=16,347 total decisions) for repeating the same arm. Mice were therefore more inclined to visit arms which were not the same arm, or immediately adjacent, and was reflected in every age group (4 months [n=58

mice], 8 months [n=46 mice], 12 months [n=31 mice], 18 months [n=22 mice]). Stability of this trait over age suggests mice remembered this strategy or immediately adopted it as a default when tested at each age point.



**Figure 6.2.** Bodyweight and open-field testing. **A.** Bodyweight of mice at each age point as mean  $\pm$  SEM (black markers and error bars) with individual animal weights (grey markers). **B.** The velocity of mice at each age point as mean  $\pm$  SEM (black markers and error bars) with individual animal velocities (grey markers). **C.** Duration of time spent in the centre (blue), inner (orange) and outer (red) regions of the arena (shown) normalised to total exploration time (mean  $\pm$  SEM). With age, mice showed increased exploration of the outer region adjacent to the wall with a reduction of time spent in the centre or inner regions.

### Mice show increased body weight and reduced locomotor function

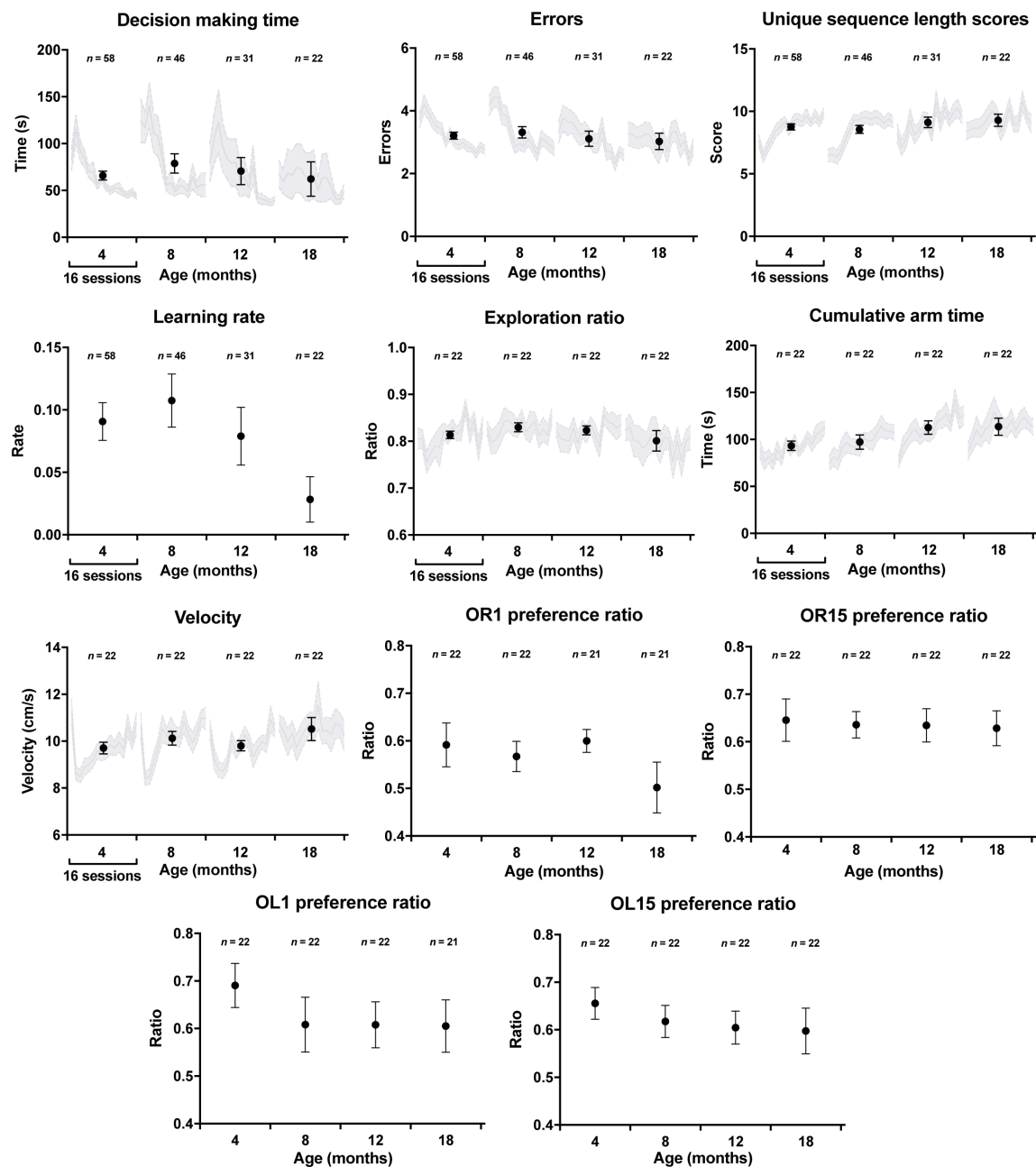
The bodyweight of mice showed monotonic increase with age (Fig. 6.2A) from a mean of  $21.51 \pm 0.25$  g (mean  $\pm$  SEM, n=22) at 4 months,  $25.70 \pm 0.53$  g (n=22) at 8 months  $29.50 \pm 0.68$  g (n=22) at 12 months and  $31.58 \pm 0.79$  g (n=22) at 18 months. Under the assumption

that the population exhibits Gaussian distribution with unequal variances across the age groups, Welch's one-way analysis of variance (ANOVA) was used to show significant changes with age ( $p < 0.0001$ ,  $n = 22$  mice), with Dunnett's corrected post hoc multiple comparisons between ages showing significant changes between 4 and 8 months of age ( $p < 0.0001$ ), 4 and 12 months of age ( $p < 0.0001$ ), 4 and 18 months of age ( $p < 0.0001$ ), 8 and 12 months of age ( $p = 0.0004$ ), and 8 and 18 months of age ( $p < 0.0001$ ).

Open-field tests (Fig. 6.2B) showed monotonic reduction in exploratory velocity in the arena with a mean velocity of  $12.71 \pm 0.44$  cm/s (mean  $\pm$  SEM,  $n = 22$ ) at 4 months,  $11.19 \pm 0.30$  cm/s ( $n = 22$ ) at 8 months,  $9.48 \pm 0.27$  cm/s ( $n = 22$ ) at 12 months and  $9.08 \pm 0.32$  cm/s ( $n = 22$ ) at 18 months. Welch's one-way ANOVA testing showed significant changes with age ( $p < 0.0001$ ), with Dunnett's corrected post hoc multiple comparisons between ages showing significant changes between 4 and 8 months of age ( $p = 0.0423$ ), 4 and 12 months of age ( $p < 0.0001$ ), 4 and 18 months of age ( $p < 0.0001$ ), 8 and 12 months of age ( $p = 0.0009$ ), and 8 and 18 months of age ( $p = 0.0002$ ).

There were also significant changes in duration spent exploring the centre, inner and outer regions of the arena normalised to the total video tracked exploration time (Fig. 6.2C). Welch's one-way ANOVA testing showed significant changes in duration spent in all of the regions with age ( $p < 0.0001$  for centre [ $n = 22$ ],  $p = 0.0001$  for inner [ $n = 22$ ],  $p < 0.0001$  for outer [ $n = 22$ ]). An approximately equal decrease in the exploration of centre and inner regions were observed with a corresponding increase from  $72.88 \pm 1.09\%$  (mean  $\pm$  SEM,  $n = 22$ ) time exploring the outer region at 4 months to  $82.26 \pm 1.05\%$  ( $n = 22$ ) at 18 months. Interestingly, the preference for exploring the centre, inner and outer regions varies depending on the time spent in the trial. A study by Shoji et al. (2016) showed more time spent in the centre region of an open-field arena in the first 5 minutes at 3 months of age, compared to 12 months of age. However, when evaluated over 120 minutes, the older mice showed an increased preference for the central region. The 10 minutes exposure time evaluated here was agreeable with the 5 minutes exposure time results presented by Shoji et al. (2016). Habituation to the environment over longer exposure times could result in changes to this exploratory preference. With an overall minimum velocity of 6.30 cm/s observed at 18 months of age, mice showed no evidence of significant locomotor limitation, which could

impact their exploratory behaviour in novel object and locations tasks. However, overall age-dependent decline does suggest an association between age and anxiety behaviours.



**Figure 6.3.** Radial maze and object recognition and location performance. Plots show daily performance (mean  $\pm$  SEM; grey solid and shaded dotted lines) across the 16 sessions alongside overall average performance for the age point (mean  $\pm$  SEM; black markers and error bars). Learning rate and object recognition and location measures show only a single representative average per age point - only one measure was taken during the object

*recognition and location tests, whilst the total 16-day session performance was required to measure learning rate. All numbers are reported for each age group for each metric plot.*

### **Population behavioural measures show increasing and decreasing trajectories**

All plots for all measured metrics can be seen in Fig. 6.3. Where possible, the raw data performance (mean  $\pm$  SEM) across the 16 days are shown in the plots with respective numbers, allowing performance across the 16-day session to be observed. Some metrics required linear regression analysis on the raw data and were plotted as a single point (mean  $\pm$  SEM). Values reported below represent the mean across the 16 days for all measures.

First, to assist in significance testing of age-related changes in performance, measures at 4 months for all metrics were averaged for each animal and tested with the D'Agostino-Pearson normality test. The 4-month data was used due to higher numbers ( $n=58$ ). Of all metrics, 6-of-11 passed the normality test, suggesting the sample population showed measurements which are consistent with sampling from Gaussian distributed population (exploration ratio,  $p=0.8754$ ; cumulative arm time,  $p=0.3027$ ; velocity,  $p=0.6759$ ; OR1,  $p=0.6358$ ; OR15,  $p=0.6020$ ; OL15,  $p=0.4067$  [ $n=58$ ]). The remaining 5-of-11 failed the normality test (decision making time,  $p<0.0001$ ; errors,  $p<0.0001$ ; unique sequence length score,  $p=0.0377$ ; OL1,  $p=0.0009$  [ $n=58$ ]). As there was no overwhelming agreement or disagreement with population normality, the population measures for all metrics were assumed to be gaussian but with non-equal variances across age groups. This is in agreement with previous behavioural studies (Shoji et al., 2016). Additionally, one-way ANOVAs do not appear to be highly sensitive to deviations from normal distributions, and as the population from normality cannot be explicitly determined, the use of a non-parametric test such as the Kruskal-Wallis test was not considered here. Accordingly, Welch's one-way ANOVA testing was used for significance measures. Despite longitudinal tracking of animals, different animal numbers for different metric measures meant no matching between measures for the same mouse was done during significance testing. This also kept the significance test identical across all measures, including those of bodyweight and open-field as presented above, even when the sample population contained longitudinally matched animals only.

Decision-making time showed a non-monotonic trajectory with a brief increase then decrease, with a mean decision-making time of  $64.06 \pm 4.71$  seconds (mean  $\pm$  SEM,  $n=58$ ) at 4 months,  $77.14 \pm 10.34$  seconds ( $n=46$ ) at 8 months,  $68.91 \pm 14.50$  seconds ( $n=31$ ) at 12 months and  $60.44 \pm 18.35$  seconds ( $n=22$ ) at 18 months. Over the 16 days, mice generally showed improvement in decision-making time, reflecting possible adaptation to the task (Fig. 6.3). At 18 months, mice were consistently faster at making decisions even at the beginning of the 16-day session suggesting retained long-term memory enabled mice to quickly adapt to the task following previous exposures at 4, 8 and 12 months of age. However, increased SEM suggests increased variability in performance. Welch's one-way ANOVA testing showed non-significant changes in performance with age ( $p=0.7009$ ), with post hoc Dunnett's multiple comparisons between ages showing no significant differences.

Errors followed a similar non-monotonic trend as decision-making time with a brief increase in error accumulation before decreasing, with a population mean error per day of  $3.22 \pm 0.11$  (mean  $\pm$  SEM,  $n=58$ ) at 4 months,  $3.33 \pm 0.18$  ( $n=46$ ) at 8 months,  $3.12 \pm 0.24$  ( $n=31$ ) at 12 months and  $3.04 \pm 0.26$  ( $n=22$ ) at 18 months. Mice showed much-improved error performance at 18 months, suggesting retained long-term memory allowed mice to adapt, as seen with decision-making time. An increased SEM at 18 months also suggests increased variability in performance between mice. Welch's one-way ANOVA testing showed no significant changes with age ( $p=0.8060$ ), with post hoc Dunnett's multiple comparisons between ages showing no significant differences.

Unique sequence length score also showed non-monotonic trajectory with a slight decrease in measure before increasing, with a score of  $8.70 \pm 0.25$  (mean  $\pm$  SEM,  $n=58$ ) at 4 months,  $8.50 \pm 0.31$  ( $n=46$ ) at 8 months,  $9.06 \pm 0.42$  ( $n=31$ ) at 12 months and  $9.22 \pm 0.49$  ( $n=22$ ) at 18 months. Compared to 4, 8 and 12 months, mice at 18 months showed the immediate capability of chaining unique arm choices when reintroduced to the radial maze (Fig. 6.3). Welch's one-way ANOVA testing showed no significant changes with age ( $p=0.5505$ ), with post hoc Dunnett's multiple comparisons between ages showing no significant differences.

Learning rate also showed non-monotonic trajectory with brief increase before decreasing with age, with a mean rate of  $0.09 \pm 0.02$  (mean  $\pm$  SEM,  $n=58$ ) at 4 months,  $0.11 \pm 0.02$  ( $n=46$ )



at 8 months,  $0.08 \pm 0.02$  (n=31) at 12 months and  $0.03 \pm 0.02$  (n=22) at 18 months. The low learning rate at 18 months is supported by a good initial error rate and consistent error performance when mice were reintroduced to the maze at 18 months. Welch's one-way ANOVA testing showed significant changes with age ( $p=0.0258$ ), with post hoc Dunnett's multiple comparisons between ages showing only a significant difference between 8 and 18 months of age ( $p=0.0368$ ).

Exploration ratio (exploration of 27 possible areas of the maze not including the centre where the mouse is introduced) again showed a brief increase in measure followed by a decrease, with a mean exploration ratio of  $0.81 \pm 0.01$  (mean  $\pm$  SEM, n=22) at 4 months,  $0.83 \pm 0.01$  (n=22) at 8 months,  $0.82 \pm 0.01$  (n=22) at 12 months and  $0.80 \pm 0.02$  (n=22) at 18 months. Note the reduced number on account of the exploration ratio coming from video tracking, which was limited to animals which underwent the complete longitudinal study from 4 months to 18 months. Across 4, 8 and 12 months, mice generally showed an increase in exploration across the 16 days, whereas, at 18 months, the exploration amount reduced. This could be mice reaching endpoint and showing reduced motor capability. Welch's one-way ANOVA testing showed no significant changes with age ( $p=0.4653$ ), with post-hoc Dunnett's multiple comparisons between ages showing no significant differences.

Cumulative arm time appeared to increase with age with a mean arm time of  $92.38 \pm 4.98$  seconds (mean  $\pm$  SEM, n=22) at 4 months,  $96.39 \pm 7.60$  seconds (n=22) at 8 months,  $111.80 \pm 7.17$  seconds (n=22) at 12 months and  $112.70 \pm 9.04$  seconds (n=22) at 18 months. Across the 16 days, mice were more inclined to spend time on the arms at the end of the session than at the beginning and was a consistent result across all age groups. Welch's one-way ANOVA testing showed no significant changes with age ( $p=0.0854$ ), with post-hoc Dunnett's multiple comparisons between ages showing no significant differences.

Velocity increased overall with a mean velocity of  $9.67 \pm 0.25$  cm/s (mean  $\pm$  SEM, n=22) at 4 months,  $10.08 \pm 0.29$  cm/s (n=22) at 8 months,  $9.77 \pm 0.22$  cm/s (n=22) at 12 months and  $10.48 \pm 0.49$  cm/s (n=22) at 18 months. Interestingly velocity appeared to consistently drop in the first few days of testing at each age point, before increasing. Re-introduction to the maze at each age point could encourage mice to immediately explore the novel environment,

with a reduction in movement once mice become more familiar with the environment. Following the establishment of a task regime after the initial days, mice could change their exploratory approach by becoming more active once again. Welch's one-way ANOVA testing showed no significant changes with age ( $p=0.4188$ ), with post-hoc Dunnett's multiple comparisons between ages showing no significant differences.

Object recognition with 1-minute retention interval measured for novel preference ratio appeared to decrease with a mean novel preference ratio of  $0.59 \pm 0.05$  (mean  $\pm$  SEM,  $n=22$ ) at 4 months,  $0.57 \pm 0.03$  ( $n=22$ ) at 8 months,  $0.60 \pm 0.02$  ( $n=22$ ) at 12 months and  $0.50 \pm 0.05$  ( $n=22$ ) at 18 months. Welch's one-way ANOVA testing showed no significant changes with age ( $p=0.4071$ ), with post-hoc Dunnett's multiple comparisons between ages showing no significant differences.

Object recognition with 15-minute retention interval measured for novel preference ratio appeared to steadily decrease with a mean novel preference ratio of  $0.65 \pm 0.04$  (mean  $\pm$  SEM,  $n=22$ ) at 4 months,  $0.64 \pm 0.03$  ( $n=22$ ) at 8 months,  $0.63 \pm 0.04$  ( $n=22$ ) at 12 months and  $0.63 \pm 0.04$  ( $n=22$ ) at 18 months. Welch's one-way ANOVA testing showed no significant changes with age ( $p=0.9935$ ), with post-hoc Dunnett's multiple comparisons between ages showing no significant differences.

Object location with 1-minute retention interval measured for novel preference ratio appeared to decrease with a mean novel preference ratio of  $0.6904 \pm 0.0465$  (mean  $\pm$  SEM,  $n=22$ ) at 4 months,  $0.6081 \pm 0.0577$  ( $n=22$ ) at 8 months,  $0.6080 \pm 0.0484$  ( $n=22$ ) at 12 months and  $0.6053 \pm 0.0552$  ( $n=22$ ) at 18 months. Welch's one-way ANOVA testing showed no significant changes with age ( $p=0.5421$ ), with post-hoc Dunnett's multiple comparisons between ages showing no significant differences.

Object location with 15-minute retention interval measured for novel preference ratio appeared to decrease with a mean novel preference ratio of  $0.6556 \pm 0.0338$  (mean  $\pm$  SEM,  $n=22$ ) at 4 months,  $0.6174 \pm 0.0338$  ( $n=22$ ) at 8 months,  $0.6044 \pm 0.0348$  ( $n=22$ ) at 12 months and  $0.5975 \pm 0.0479$  ( $n=22$ ) at 18 months. Welch's one-way ANOVA testing showed no

significant changes with age ( $p=0.6834$ ), with post-hoc Dunnett's multiple comparisons between ages showing no significant differences.

### **Population behaviour shows minimal variability and skewness changes**

Analysis of population skewness (Supplementary Fig. 6.1) revealed that some metrics tended to increase skewness with age, such as decision-making time and errors. This suggests a non-consistent trajectory of individual cognitive performance, with subsets of the population improving or declining in some measures. However, linear regression analysis on the population skew revealed no significant correlation for all metrics ( $p$ -values reported on plots of Supplementary Fig. 6.1). Linear regression was performed by taking the midpoint of each age group in days (P133 at 4 months, P265 at 8 months, P393 at 12 months and P574 at 18 months) to compensate for non-uniform intervals between testing ages (4 months between the first three age points and 6 months for the last).

Analysis of the coefficient of variability (CV; Supplementary Fig. 6.2) of the population revealed the variability of population performance with respect to the mean. Some metrics appeared to show an increase in CV with age, such as decision-making time and learning rate. However, linear regression analysis on population CV across all ages showed only decision-making time increased significantly ( $R=0.9985$ ,  $R^2=0.9971$ ,  $p<0.025$ ,  $n=58$ , 46, 31, 22 at 4, 8, 12 and 18 months of age), with an increase in CV from 0.5593 at 4 months to 1.4242 at 18 months.

### **Behavioural measures can show stable trajectory with age for individual animals**

Individual metric measures for the radial maze task, object recognition and location tasks cover different cognitive attributes and have different units. Directly comparing inter-animal and intra-animal performance across the different age points and comparing between metrics is, therefore, not trivial. This was addressed by calculating the Z-score of each animal for each metric, allowing performances to be standardised with respect to the cohort. Standardised Z-scores allowed metrics to be combined and directly compared, and also tracked across ages without worrying about compounding issues with overall population changes. In other words, the Z-scores can be used to address whether mice which performed well at earlier ages, also performed well at later ages with respect to the sample population mean. Metric measures

could also be combined as the measures were standardised. The open-field measures were omitted from this analysis as the task does not measure any behavioural traits and was only used to assess whether mice showed limited motor movement prior to both novel object tasks.

For each metric, an animal's Z-score was calculated as the animal performance,  $x$ , minus the population mean,  $\mu$ , divided by the standard deviation,  $\sigma$ .

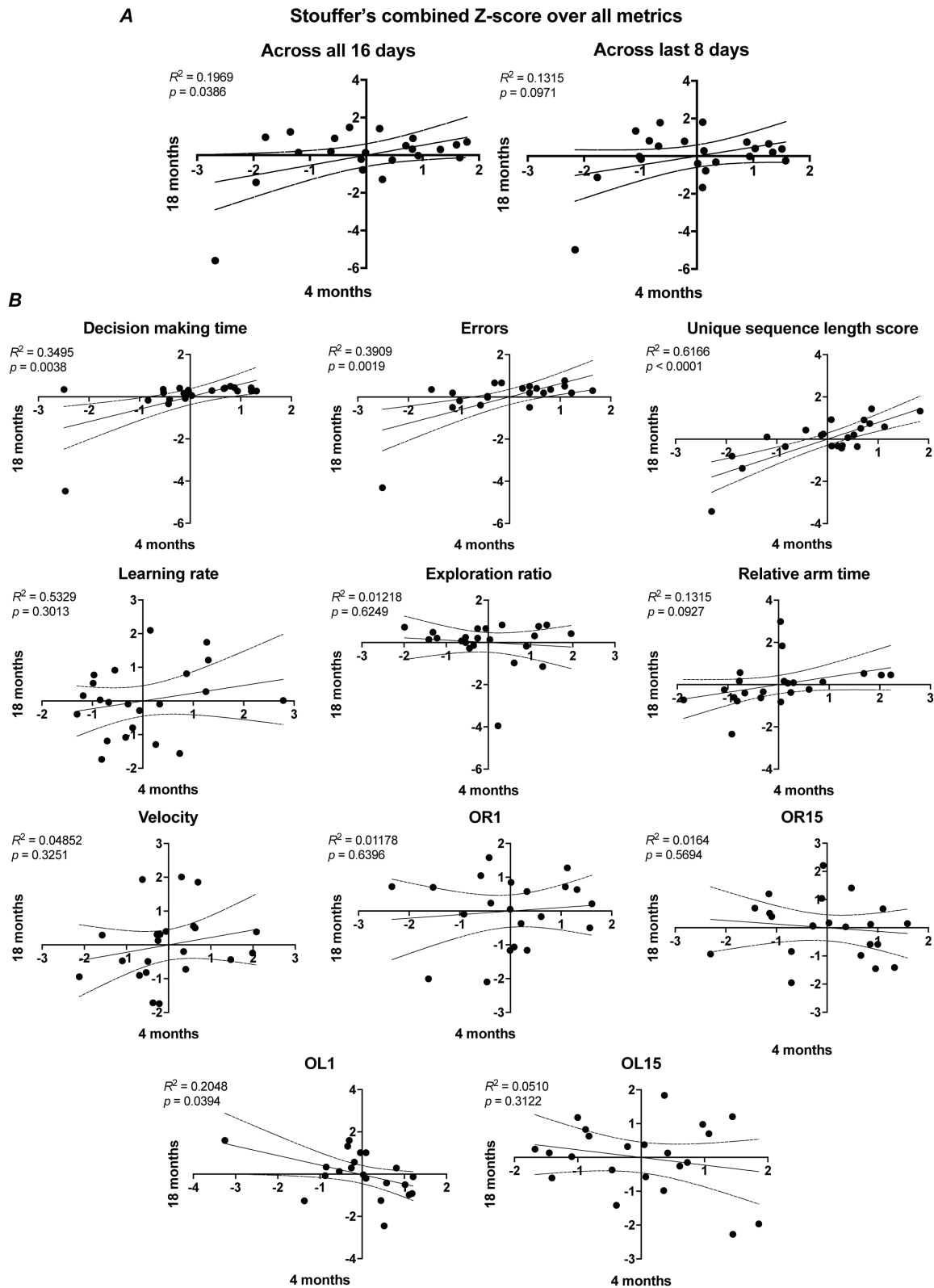
$$Z_i = \frac{x - \mu}{\sigma} \quad (\text{Eq. 6.5})$$

Individual animals that performed exceptionally well or poorly in the metric had a more positive or negative Z-score. Scores for some metrics were inverted to correlate a higher Z-score value with better performance as detailed in 'Methods'.

When mice were introduced into the maze at each RM age point, it was possible mice had forgotten the environment and task and required several days of familiarisation. We, therefore, wanted to assess whether performance across all 16 days had higher predictive power than the performance over the last 8 days, where mice might be more familiar with the task. All Z-scores for each radial maze metric plus object location and recognition scores, were combined into a single representative Z-score using the Stouffer's method below.

$$Z = \frac{\sum_i^k Z_i}{\sqrt{k}} \quad (\text{Eq. 6.6})$$

$Z_i$  represents the Z-score for metric  $i$ , and  $k$  represent the number of metrics. Over all 16 days, there was a higher coefficient of determination for performance at 18 months given animal performance at 4 months (Fig. 6.4A,  $R^2=0.1969$ ,  $p=0.0386$ ,  $n=58$ ) than the same evaluation when using the last 8 days of measures in the radial maze (Fig. 6.4A,  $R^2=0.1315$ ,  $p=0.0971$ ,  $n=58$ ). Measures across all 16 days, therefore, provided more predictive power and were therefore used for the rest of the analysis.



**Figure 6.4.** Tracking individual animal performance with age. **A.** Evaluating Stouffer's combined metric Z-score for all metrics using all 16 days and the last 8 days of radial maze task ( $n=22$  mice). **B.** Z-score performance for each animal for each metric at 18 months versus

*performance at 4 months. Linear regression coefficient of determination  $R^2$  and p-value are reported on each plot (n=22 mice).*

To address whether individual mouse performance showed predictable linear trajectories for each metric, linear regression was used on performance at 18 months against 4 months (Fig. 6.4B). Decision making time ( $R^2=0.3495$ ,  $p=0.0038$ ,  $n=22$ ), errors ( $R^2=0.3909$ ,  $p=0.0019$ ,  $n=22$ ) and unique sequence length score ( $R^2=0.6166$ ,  $p<0.0001$ ,  $n=22$ ) showed significant positive correlations. Interestingly, OL1 showed a significant negative correlation ( $R^2=0.2048$ ,  $p<0.0394$ ,  $n=22$ ), suggesting high early life high performance led to later life poor performance. All other metrics showed no significant increase or decrease in performance between 18 months and 4 months performance.

To assess whether individual mouse performance trajectory was consistent across each possible time point comparison (six comparisons in total for 8 versus 4 months, 12 versus 4 months, 18 versus 4 months, 12 versus 8 months, 18 versus 8 months and 18 versus 12 months), the  $R^2$  was averaged across all age comparisons and plotted in order of high-to-low for each metric (Fig. 6.5A). A general low coefficient of determination (mean  $R^2<0.4$ ,  $n=22$ ) between individual animal performances for each metric suggests mediocre predictability of late age performance given early age performance. Object recognition and location showed the worst coefficient of determination (mean  $R^2<0.1$ ,  $n=22$ ), whilst unique sequence length score proved to be the most predictive in agreement with the 18 versus 4 months performance comparison alone (Fig. 6.4B).

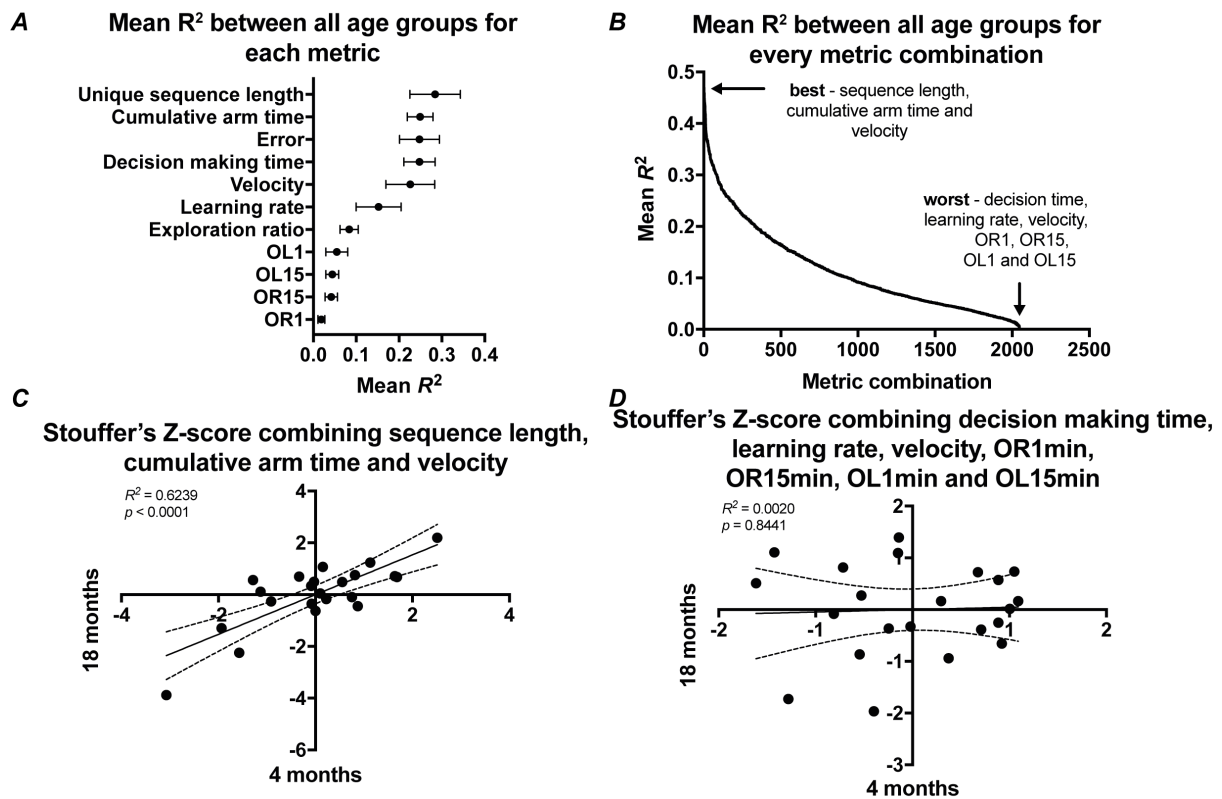
Although open-field testing was not used to evaluate cognitive behaviour, to see whether open-field velocity showed stability across age, linear regression was performed on performance at 18 months against 4 months showing low and insignificant linear trajectory and predictability of late-life velocity measures given early life measures ( $R=0.3160$ ,  $R^2=0.0999$ ,  $p=0.1519$ ,  $n=22$ ). When evaluated across all six time point comparisons, open-field velocity gave an improved mean coefficient of determination over the object location and recognition measures and exploration ratio (mean  $R^2=0.14$ ), but was still weak overall. This suggests that a task not designed for cognitive behavioural testing does not give particularly

stable measures of performance on an individual animal basis and reinforces the experimental design and significance of behavioural test design.

### **Combining metrics to improve individual animal tracking**

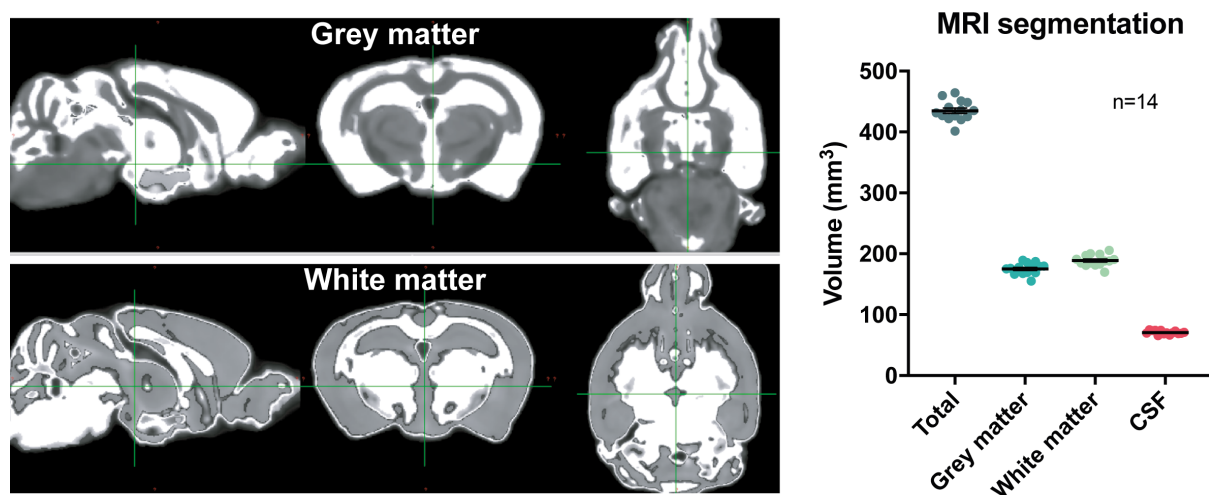
To address whether performance metrics could be combined into a more predictive trajectory across ageing, the Z-scores for each animal were combined using all possible inclusive and exclusive combinations of metrics (a total of 2,047, including single metrics). Z-scores were combined using Stouffer's method. The highest correlated performance metric combination assessed by the coefficient of determination across all age groups (Fig. 6.5B), was a combination of sequence length, cumulative arm time and velocity ( $R^2=0.48$ ,  $n=22$ ). The worst possible combination was decision-making time, learning rate, velocity, OR1, OR15, OL1 and OL15 ( $R^2<0.01$ ,  $n=22$ ). Mice in 1- and 15-minute object location tests performed better than the object recognition tests, suggesting novel object location measures are more reliable than novel object recognition (Benice et al., 2006).

Plotting individual animals using a combined Z-score of sequence length, cumulative arm time and velocity for 18 versus 4 months performance yielded a statistically significant non-zero correlation ( $R=0.7899$ ,  $R^2=0.6239$ ,  $p<0.0001$ ,  $n=22$ ). The high correlation suggests this combination of sequence length, cumulative arm time and velocity performance could show stable trajectory throughout ageing with a high level of predictability from early age performance for individual animals (Fig. 6.5C). The predictability using the combined metrics was comparable to using only the unique sequence length score for 18 versus 4 months performance ( $R^2=0.6166$ ,  $p<0.0001$ ,  $n=22$ ; Fig. 6.4B). In comparison, the worst combination of decision-making time, learning rate, velocity, OR1, OR15, OL1 and OL15 for 18 versus 4 months yielded a statistically non-significant correlation (Fig. 6.5D,  $R=0.0447$ ,  $R^2=0.0020$ ,  $p=0.8441$ ,  $n=22$ ).



**Figure 6.5.** Combining performance metrics for individual animal performance using standardised Z-score to reflect individual animal performance. **A.** All age groups were compared amongst themselves for each metric with linear regression modelling to determine the correlation and predictability in individual animal performance. Taking the mean and SEM, metrics were ordered by their mean coefficient of determination ( $R^2$ ), representing how well a linear correlation describes the performance of an individual animal at one age point given its performance at a younger age. **B.** All possible combinations of metrics were evaluated using the same linear regression approach and ordered by overall mean  $R^2$ . Combining metrics was achieved by combining the Z-scores for each animal using Stouffer's method. A combination of unique sequence length, cumulative arm time and velocity gave the best overall linear correlation between all age groups ( $R^2=0.4761$ ,  $n=22$ ), whilst decision-making time, learning rate, velocity, OR1, OR15, OL1 and OL15 yielded the worst combination ( $R^2=0.0035$ ,  $n=22$ ). **C.** A plot of individual animal performance for the combined metric of unique sequence length, cumulative arm time and velocity at 18 months against performance at 4 months ( $R=0.7899$ ,  $R^2=0.6239$ ,  $p<0.0001$ ,  $n=22$ ), showing high correlation and levels of predictability. **D.** A plot of individual animal performance for the combined metric of decision-making time, learning rate, velocity, OR1, OR15, OL1 and OL15 at 18 months against performance at 4 months ( $R=0.0447$ ,  $R^2=0.0020$ ,  $p=0.8441$ ,  $n=22$ ), showing zero correlation and predictability.





**Figure 6.6.** Grey matter and white matter segmentation from MRI T1 scans (left) using FSL. Plots (mean  $\pm$  SEM) of segmentation for white matter, grey matter, cerebrospinal fluid (CSF) and total volume (n=14).

#### MRI allows for measuring individual brain volumes

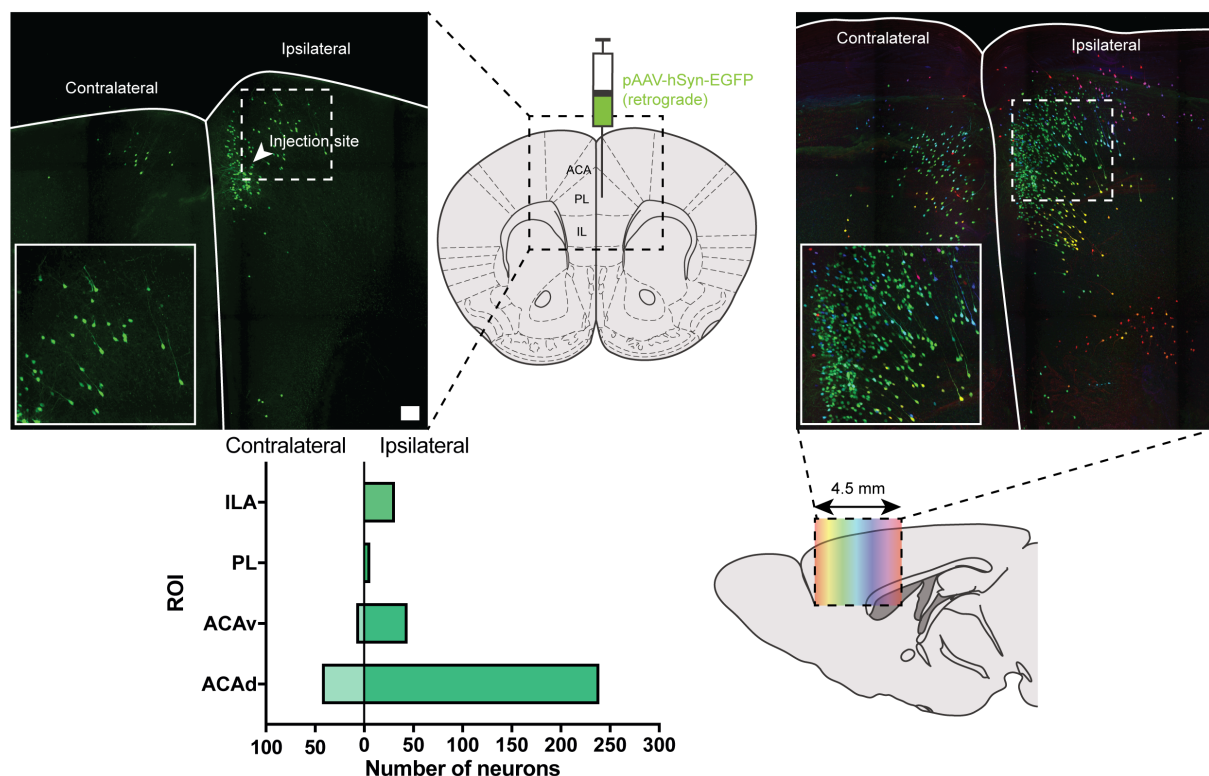
Using FSL, white matter, grey matter and cerebrospinal fluid (CSF) volumes were segmented following T1 MRI after behavioural testing at 18 months (Fig. 6.6, n=14). Mean total brain volume was  $435.0 \pm 4.4 \text{ mm}^3$  (mean  $\pm$  SEM, n=14), with a higher white matter content at  $189.0 \pm 2.5 \text{ mm}^3$  (n=14) compared to grey matter at  $175.3 \pm 2.5 \text{ mm}^3$  (n=14). T2 and DTI scan data also exists, but is not shown here. DTI would be particularly valuable for measuring fractional anisotropy and white matter tract degradation.

#### STP imaging could reveal single-cell resolution anatomical correlates of cognitive ageing

Retrograde labelling in the PFC using pAAV-hSyn-EGFP and high-resolution STP imaging across the brain allowed projection neurons from contralateral PFC to be identified (n=1 mouse). Projection axons establish inter-hemispheric connectivity in the PFC through the corpus callosum, with axons appearing to extend as far as 4.5 mm posterior to the injection site as they projected across corpus callosum (Fig. 6.7). Projection neurons on the contralateral PFC categorise as IT neurons which establish bilateral cortical connectivity (Dembrow & Johnston, 2014).

Delineation of the PFC into infralimbic (IL), prelimbic (PL) and dorsal- and ventral- anterior cingulate area (dACA, vACA) regions using SimpleElastix (see 'Chapter 3') in combination with manual counting yielded a total of 320 neurons labelled on the ipsilateral side, and a total of 51 labelled IT neurons projecting from the contralateral PFC. The majority of ipsilateral pyramidal neurons came from ACAd (239 neurons) followed by ACAv (44 neurons), ILA (31 neurons) and last PL (6 neurons) (Fig. 6.7). Distribution of labelled pyramidal neurons in the contralateral PFC was similar to the ipsilateral distribution with the majority coming from ACAd (43 neurons) and ACAc (8 neurons) with none seen in the PL or ILA regions. Numbers of IT projection neurons in the PFC could potentially correlate with levels of cognitive ability in aged mice (Cabeza, 2002)(see 'Discussion' and 'Future Perspectives').

Neuronal tracing of a small subset of pyramidal neurons across layer 2/3, 5 and 6a of the contralateral PFC show three different pyramidal neuron classes. The layer 2/3 neuron showed stereotypical type I morphology, with dendrites and collaterals spanning layer 2/3 and 5 (Larsen & Callaway, 2006). The layer 5 neuron showed tufted morphology with distal bifurcation along the apical dendrite in the superficial layers of the cortex (Yun Wang et al., 2018). Lastly, the layer 6 neuron showed narrow tufted morphology (Yun Wang et al., 2018).



**Figure 6.7.** Demonstration of high-resolution connectivity mapping in the PFC following pAAV-hSyn-EGFP (retrograde) virus injection. Retrograde labelling of ipsilateral and intratelencephalic projection neurons from the contralateral PFC reveals inter-hemispheric connectivity (top-left). Depth-coded Z-stack (right, colour coding as indicated in the key) across a 4.5 mm thick volume of the PFC shows the extent of projection neuron labelling. Axons from contralateral PFC appear to extend as far as 4.5 mm posterior to the injection site when projecting across the corpus callosum. Manual quantification of automatically delineated regions (using SimpleElastix) of ILA, PL, ACAv and ACAd show majority ipsilateral and contralateral labelling in the ACAd and ACAv, followed by IL and PL. Low correspondence between labelled neuron number and injection target of layer 2/3 PL could reflect errors in stereotaxic injection or automated registration procedure. Infralimbic, IL; prelimbic, PL; dorsal and ventral anterior cingulate area, dACA/vACA.

## 6.4 DISCUSSION

Here, age-related changes in cognitive behaviour with a nine-arm radial maze and novel object location and recognition were evaluated using multiple extracted measures; arm decision-making time, errors, unique sequence length score, learning rate, exploration ratio, cumulative arm time, velocity and novel object preference ratio for 1- and 15-minute retention intervals. Throughout the 16-day session with the radial maze, mice adapted to the task showing improvement in all metrics such as time between decisions, the number of errors they made (classified as missed arms) and the number of unique arm explorations they accomplished. Even in late life at 18 months of age, mice retained cognitive performance by displaying little impact on overall population measures. At each re-exposure to the maze at 8, 12 and 18 months, mice retained high levels of performance. Improvement across the 16-days of testing was also observed, showing that mice were adapting to the task. These results could be reflective of results seen in longitudinal studies with humans (Salthouse, 2014), and a demonstration of long term memory retention in mice.

Whilst non-significant, some metrics such as decision-making time, errors and unique sequence length score appeared to show increased variability with age as assessed by population skewness and CV. This meant that although the overall population performance could be unchanged, individual mice tended to non-uniformly improve or decline in specific measures. Increased numbers within the population of tested animals would improve the reliability of this observation.

Surprisingly, despite older age, mice at 18 months appeared to show higher average velocity in the task compared to 4, 8 and 12 months. This could reflect long-term memory of the task, higher familiarity with the task, and reduced anxiety therein. Contrary to this, the open-field arena velocity decreased significantly with age. The open-field arena was evaluated over three days for which the first few days of RM testing showed reduced velocity before increasing for the remainder of the testing session. This could suggest that exploratory behaviour in mice following re-exposure into the radial maze was similar to the open-field arena. Once the mice establish the task regime in the radial maze, they increase exploratory behaviour shown by increased velocity, but this was missed in the open-field arena because three days exposure was too short for task adaptation. A study in mice and rats showed that

rodents acquired cumulative learning of a task such as to distinguish between novel objects in a familiar location from novel objects in a novel location (Genzel et al., 2019). Repeat exposure and spatial memory, therefore, suggests some degree of long-term memory accumulation that affected task performance when mice were reintroduced into the maze.

Novel object preference during novel object recognition at 1- and 15- minute retention intervals did not appear to show any difference with age. However, whilst not significant, novel object location showed a gradual decline, in agreement with previous observations (Benice et al., 2006). Novel object location tests could, therefore, be more reliable as a behavioural test compared to novel object recognition.

With regards to tracking and correlating individual animal performance, unique arm sequence score showed the highest level of correlation between all possible age group comparisons when evaluating data across all 16 days and gave good predictability when only looking at 18 versus 4 months performance ( $R=0.7852$ ,  $R^2=0.6166$ ,  $p<0.0001$ ,  $n=22$ ). Mice, therefore, tended to more uniquely, and less stochastically, explore the arms of the maze and continued to do so in later life. This was followed by cumulative arm time, error, decision-making time and velocity, in order of coefficient of determination. Learning rate, exploration ratio and all the object location and recognition tasks showed the lowest predictive power. Slight improvement to predictability was found by combining a variety of metrics out of 2,047 possible combinations. A combination of unique sequence length score, cumulative arm time and velocity gave the highest level of correlation ( $R=0.7899$ ,  $R^2=0.6239$ ,  $p<0.0001$ ,  $n=22$ ) for 18 versus 4 months performance suggesting that a combination of these metrics could be used to predict performance in later life. However, this is directly comparable with the unique sequence length score alone for 18 versus 4 months performance ( $R=0.7852$ ,  $R^2=0.6166$ ,  $p<0.0001$ ,  $n=22$ ). Whilst combining metrics improved predictability, it provided a small improvement overall. Furthermore, cumulative arm time could be negatively correlated, although not perfectly, with velocity measures leading to the small improvement in predictability. Again, the object location and recognition measures proved to be the least correlative, especially when combined with decision-making time, learning rate and velocity. Low agreement between object location and recognition scores could come down to a lack of repeated measurements. However, the spontaneous recognition of novel objects forms

the crux of the task which becomes lost with repeated measurements on the same animal (Blaser & Heyser, 2015). High levels of variability between spontaneous inter-animal decisions could, therefore, be ultimately responsible for the low correlation scores for those four metrics.

Overall, it appeared that individual mice showed some degree of predictive linear performance trajectory depending on the measures being considered, such as unique sequence length score and decision-making time. This result could help predict mice which were on-path to late-life cognitive decline and assist in treatment or intervention design. For example, Tarantini et al. (2019), have shown that cerebral blood flow induced cognitive decline in aged mice can be treated with nicotinamide mononucleotide (NMN) which improves neurovascular coupling. Diet induced cognitive decline can also be mediated with regular exercise sessions (Graham et al., 2019).

However, mice that performed well at 4 months of age and could be considered in the top percentiles of the population, did not necessarily perform well at later time points leading to low coefficients of determination when comparing between age points. Mice also did not reliably perform over the 16-day trial with the radial maze. As is the challenge with many behavioural tasks using animal models, cognitive ability is inferred from task performance which is inferred from the behaviour of animals. Without a way to communicate the task to mice, there are limitations in how much can be gleaned from their task performance. Mice did not show stable performance or explorative behaviour across the 16 days, demonstrating why all measures needed to be averaged over all 16 days rather than a subset of days. In addition, the 16-day session in the radial maze allowed mice to learn the task, which was not possible with either of the object recognition or locations tests. Behavioural measures from a learnable task with repeated exposures at older age points could show a different trajectory of cognitive performance compared to single exposure tasks.

Alongside behavioural data, it was also possible to take anatomical measures. As evidenced with an *ex-vivo* study, grey matter, white matter and CSF volume were measured for a subset of mice that survived the total 18 month test and were imaged with MRI. Using an MRI atlas and registration (as in 'Chapter 3'), volumes for different brain regions could be measured

and compared across age. T1, T2 and DTI scans could also lead to studies on fractional anisotropy and WMH prevalence. MRI is also non-invasive meaning such a technique could be performed *in-vivo* at each age point. MRI studies in living mice are also beneficial over the *ex-vivo* case where data can be systematically affected by changes to brain structure following brain extraction and fixation methods.

Although not achievable through an *ex-vivo* approach, retrograde viral injection of IT projection neurons in the PFC combined with STP imaging could reveal levels of local and bilateral neuronal connectivity (Fig. 6.7). Retrograde labelling showed remarkable efficacy and clear labelling across the hemisphere, allowing three classifications of IT neurons to be morphologically identified. Classes of IT projection neurons and the number of labelled contralateral and ipsilateral neurons could be correlated with behavioural data and with observations of bilateral PFC recruitment in human subjects (Cabeza, 2002). Synaptic labelling techniques such as mGRASP (J. Kim et al., 2011) could also reveal synaptic connectivity levels and help address what changes and compensatory mechanisms might be deployed in mice such that they maintain apparent high levels of cognitive function even up to their endpoint.

Ultimately, the work presented here demonstrates a novel exploration of behavioural changes with age, how individual may be tracked with age, and how cognitive ageing changes could be correlated with anatomical changes in the brain. Some of the experimental work demonstrated in this chapter was impacted by COVID-19. For example, STP imaging of brain samples which had been imaged with MRI was not feasible. However, this would help validate the anatomical measures of low- and high-resolution whole-brain imaging techniques. In addition, further experiments for retrograde labelling of IT projection neurons in the PFC could not be completed. Some of these opportunities are explored under 'Future Directions' below.

## 6.5 FUTURE DIRECTIONS

Recording large numbers (n=58) of mice in the radial maze over 16 days at 4, 8, 12 and 18-month time points, yielded a large amount of data, which was split across manual event logging, video data, and subsequent video tracking data. Here, decision-making time, errors, unique arms sequence length score, learning rate, exploration ratio, cumulative arm time and velocity were extracted from manual event logs and video tracking using EthoVision software. However, there are several additional measures which could be extracted from the radial maze data, as explained below.

One measure could assess frequency of food consumption on the arms. In general, mice preferred to consume the food pellets later on in the 16-day trial and more so in later life (data not shown here). Additionally, it would be possible to measure the number of instances where mice explored the full extent of the arm rather than exploring only the area adjacent to the bridge. Another measure could look at the weight of animals during testing and whether this correlated with food consumption and explorative behaviour in the radial maze task. Additionally, video tracking using software such as DeepLabCut (Mathis et al., 2018) could help classify posture, pose and solitary animal behaviour (Lezak et al., 2017). Do mice show different levels of grooming and rearing when exploring the maze, for example, and could maze performance be strongly correlated with anxiety level as inferred from posture analysis? Additionally, when exploring the maze, how stochastic or deliberate were exploration paths? Unsurprisingly, mice tended to spend the majority of their time in the centre of the maze and the bridges. However, this measure could assess whether they were erratically exploring or showing limited movement. Visual cues posted on the end of each arm were chosen to provide a spatial reference for the mice. However, mice could only observe the cues if the bridges were explored. Therefore, one measure could assess how frequently the bridges were being explored, and how many explored bridges led to an immediate exploration of the arm.

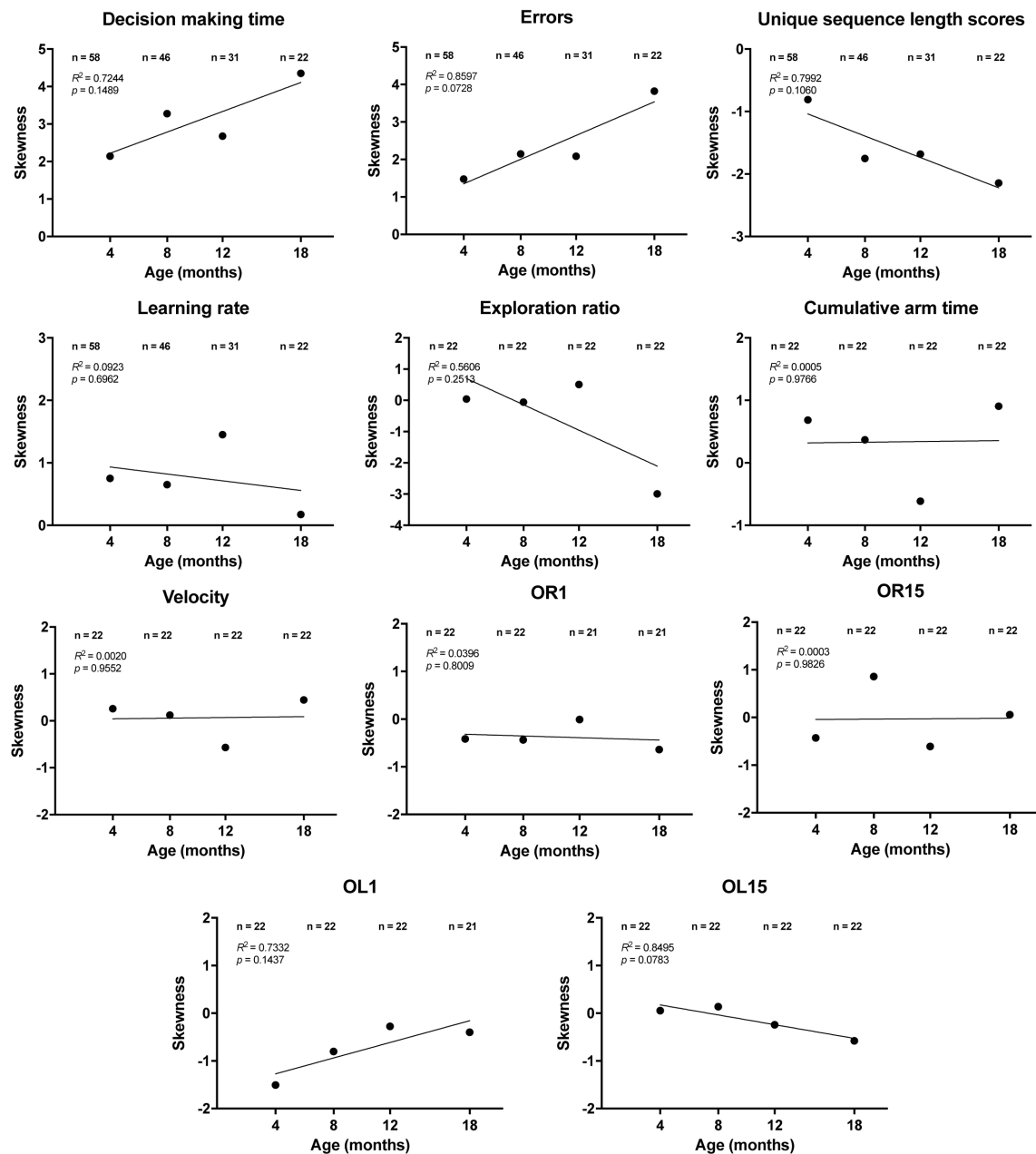
An exciting extension to the study would be to combine structural data using MRI and STP imaging, with behavioural data. The experimental basis of a study like this is briefly discussed in 'Results'. MRI imaging of white/grey matter volumes in subregions of the brain and measurements of fractional anisotropy could reveal interesting correlates with cognitive



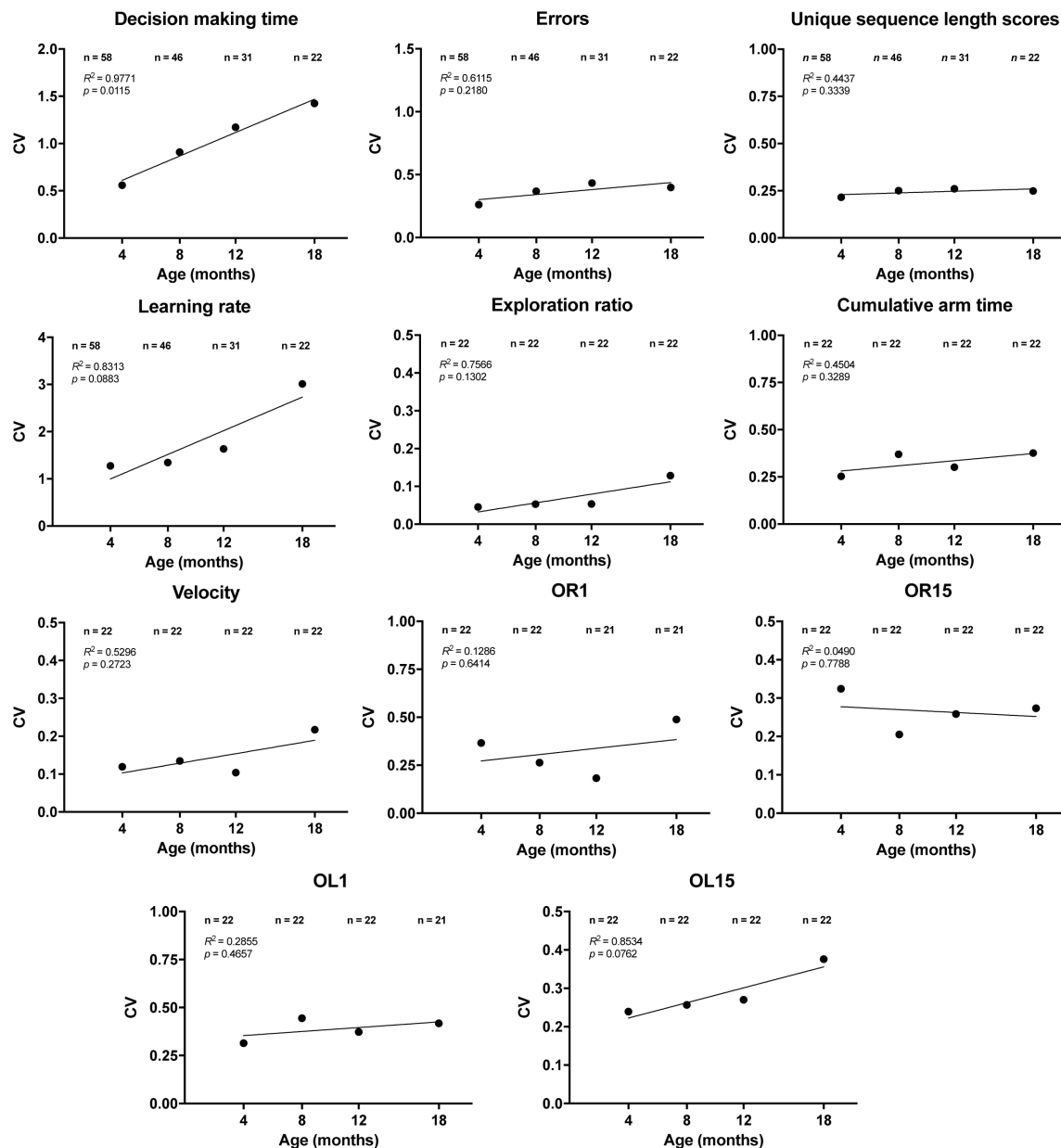
performance. One study could pair MRI imaging with high-resolution STP imaging to compare results from segmentation. Do segmentation and brain volumes measured from MRI data agree with measures from high-resolution STP imaging?

In a new comprehensive study, mice could be subjected to behavioural tests such as the radial maze and imaged with MRI to measure brain volume and fractional anisotropy across age. Select mice could also be injected with a retrograde virus or mGRASP and imaged after behavioural task completion to measure levels of connectivity in the brain, either through numbers of projection neurons or synapse numbers.

## 6.6 SUPPLEMENTARY DATA



**Supplementary Figure 6.1.** Skewness measurements for each radial maze and object recognition and location task for the population. Skewness values represents the mean across all animals over the 16 trial days. For accurate linear regression modelling, the midpoint of each age group (in days) was used (P133 at 4 months, P265 at 8 months, P393 at 12 months and P574 at 18 months) to compensate for non-uniform intervals between testing ages (4 months between the first three age groups and 6 months for the last). Numbers,  $R^2$  and p-value, are reported on the plots. None of the metrics showed significant changes with skewness during ageing.



**Supplementary Figure 6.2.** Coefficient of variation (CV) measurements for each radial maze and object recognition and location task for the population. CV value represents the variability of the mean performance of each animal over the 16 trial days. For accurate linear regression modelling, the midpoint of each age group (in days) was used (P133 at 4 months, P265 at 8 months, P393 at 12 months and P574 at 18 months) to compensate for non-uniform intervals between testing ages (4 months between the first three age groups and 6 months for the last). Numbers,  $R^2$  and  $p$ -value, are reported on the plots. Only decision-making time showed a significant change in CV, with an increase from 0.5 to 1.5 over the 4 to 18 months of testing.

## REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Yangqing, J., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*.  
<https://www.tensorflow.org/>
- Abraham, N., & Khan, N. M. (2018). A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation. *Proceedings - International Symposium on Biomedical Imaging, 2019-April*, 683–687. <http://arxiv.org/abs/1810.07842>
- Abuhamdah, R. M., Hussain, M. D., Chazot, P. L., & Ennaceur, A. (2016). Pre-training in a radial arm maze abolished anxiety and impaired habituation in C57BL6/J mice treated with dizocilpine. *Physiology and Behavior*, 164, 353–360.  
<https://doi.org/10.1016/j.physbeh.2016.06.017>
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., ... Craig Venter, J. (2000). The genome sequence of *Drosophila melanogaster*. In *Science* (Vol. 287, Issue 5461, pp. 2185–2195). American Association for the Advancement of Science. <https://doi.org/10.1126/science.287.5461.2185>
- Allen, J. S., Bruss, J., Brown, C. K., & Damasio, H. (2005). Normal neuroanatomical variation due to age: the major lobes and a parcellation of the temporal region. *Neurobiology of Aging*, 26(9), 1245–1282. <https://doi.org/10.1016/j.neurobiolaging.2005.05.023>
- Amato, S. P., Pan, F., Schwartz, J., & Ragan, T. M. (2016). Whole brain imaging with serial two-photon tomography. *Frontiers in Neuroanatomy*, 10, 31.  
<https://doi.org/10.3389/fnana.2016.00031>
- Andel, R., Vigen, C., Mack, W. J., Clark, L. J., & Gatz, M. (2006). The effect of education and occupational complexity on rate of cognitive decline in Alzheimer's patients. *Journal of the International Neuropsychological Society*, 12(1), 147–152.  
<https://doi.org/10.1017/S1355617706060206>
- Angeli, C. A., Boakye, M., Morton, R. A., Vogt, J., Benton, K., Chen, Y., Ferreira, C. K., & Harkema, S. J. (2018). Recovery of Over-Ground Walking after Chronic Motor Complete Spinal Cord Injury. *New England Journal of Medicine*, 379(13), 1244–1250.

<https://doi.org/10.1056/NEJMoa1803588>

- Arcelli, P., Frassoni, C., Regondi, M. C., Biasi, S. De, & Spreafico, R. (1997). GABAergic neurons in mammalian thalamus: A marker of thalamic complexity? *Brain Research Bulletin*, 42(1), 27–37. [https://doi.org/10.1016/S0361-9230\(96\)00107-4](https://doi.org/10.1016/S0361-9230(96)00107-4)
- Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., & Gee, J. C. (2014). The Insight ToolKit image registration framework. *Frontiers in Neuroinformatics*, 8(APR). <https://doi.org/10.3389/fninf.2014.00044>
- Azevedo, A. S., Grotek, B., Jacinto, A., Weidinger, G., & Saúde, L. (2011). The regenerative capacity of the zebrafish caudal fin is not affected by repeated amputations. *PloS One*, 6(7), e22820–e22820. <https://doi.org/10.1371/journal.pone.0022820>
- Azevedo, F. A. C. C., Carvalho, L. R. B. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L. L., Leite, R. E. P. P., Filho, W. J., Lent, R., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5), 532–541. <https://doi.org/10.1002/cne.21974>
- Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., & Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. *Nature*, 529(7586), 345–350. <https://doi.org/10.1038/nature16468>
- Baloch, S., Verma, R., Huang, H., Khurd, P., Clark, S., Yarowsky, P., Abel, T., Mori, S., & Davatzikos, C. (2008). Quantification of Brain Maturation and Growth Patterns in C57BL/6J Mice via Computational Neuroanatomy of Diffusion Tensor Images. *Cerebral Cortex*, 19(3), 675–687. <https://doi.org/10.1093/cercor/bhn112>
- Bartzokis, G., Beckson, M., Lu, P. H., Nuechterlein, K. H., Edwards, N., & Mintz, J. (2001). Age-related changes in frontal and temporal lobe volumes in men: a magnetic resonance imaging study. *Archives of General Psychiatry*, 58(5), 461–465. <http://www.ncbi.nlm.nih.gov/pubmed/11343525>
- Benice, T. S., Rizk, A., Kohama, S., Pfankuch, T., & Raber, J. (2006). Sex-differences in age-related cognitive decline in C57BL/6J mice associated with increased brain microtubule-associated protein 2 and synaptophysin immunoreactivity. *Neuroscience*, 137(2), 413–423. <https://doi.org/10.1016/j.neuroscience.2005.08.029>
- Bennett, C. M., Miller, M. B., & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple

- comparisons correction. *NeuroImage*, 47, S125. [https://doi.org/10.1016/s1053-8119\(09\)71202-9](https://doi.org/10.1016/s1053-8119(09)71202-9)
- Berezin, M. Y., & Achilefu, S. (2010). Fluorescence lifetime measurements and biological imaging. *Chemical Reviews*, 110(5), 2641–2684. <https://doi.org/10.1021/cr900343z>
- Bickford, M. E., Slusarczyk, A., Dilger, E. K., Krahe, T. E., Kucuk, C., & Guido, W. (2010). Synaptic development of the mouse dorsal lateral geniculate nucleus. *Journal of Comparative Neurology*, 518(5), 622–635. <https://doi.org/10.1002/cne.22223>
- Bickford, M. E., Wei, H., Eisenback, M. A., Chomsung, R. D., Slusarczyk, A. S., & Dankowski, A. B. (2008). Synaptic organization of thalamocortical axon collaterals in the perigeniculate nucleus and dorsal lateral geniculate nucleus. *The Journal of Comparative Neurology*, 508(2), 264–285. <https://doi.org/10.1002/cne.21671>
- Bickford, M. E., Zhou, N., Krahe, T. E., Govindaiah, G., & Guido, W. (2015). Retinal and tectal “Driver-Like” inputs converge in the shell of the mouse dorsal lateral geniculate nucleus. *Journal of Neuroscience*, 35(29), 10523–10534. <https://doi.org/10.1523/JNEUROSCI.3375-14.2015>
- Bink, D. I. (2016). *Neuropathological changes in mouse models of cardiovascular diseases*.
- Blaser, R., & Heyser, C. (2015). Spontaneous object recognition: A promising approach to the comparative study of memory. *Frontiers in Behavioral Neuroscience*, 9.
- Bloice, M. D., Stocker, C., & Holinger, A. (2017). Augmentor: An Image Augmentation Library for Machine Learning. *The Journal of Open Source Software*, 2(19), 432. <https://doi.org/10.21105/joss.00432>
- Born, G., & Schmidt, M. (2007). GABAergic pathways in the rat subcortical visual system: a comparative study in vivo and in vitro. *European Journal of Neuroscience*, 26(5), 1183–1192. <https://doi.org/10.1111/j.1460-9568.2007.05700.x>
- Bresadola, M. (1998). Medicine and science in the life of Luigi Galvani (1737-1798). *Brain Research Bulletin*, 46(5), 367–380. [https://doi.org/10.1016/S0361-9230\(98\)00023-9](https://doi.org/10.1016/S0361-9230(98)00023-9)
- Brickman, A. M., Siedlecki, K. L., Muraskin, J., Manly, J. J., Luchsinger, J. A., Yeung, L. K., Brown, T. R., DeCarli, C., & Stern, Y. (2011). White matter hyperintensities and cognition: Testing the reserve hypothesis. *Neurobiology of Aging*, 32(9), 1588–1598. <https://doi.org/10.1016/j.neurobiolaging.2009.10.013>
- Briggman, K. L., Helmstaedter, M., & Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, 471(7337), 183–190.

<https://doi.org/10.1038/nature09818>

- Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and Aging*, 17(1), 85–100. <http://www.ncbi.nlm.nih.gov/pubmed/11931290>
- Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging gracefully: Compensatory brain activity in high-performing older adults. *NeuroImage*, 17(3), 1394–1402. <https://doi.org/10.1006/nimg.2002.1280>
- Cabeza, R., Daselaar, S. M., Dolcos, F., Prince, S. E., Budde, M., & Nyberg, L. (2004). Task-independent and task-specific age effects on brain activity during working memory, visual attention and episodic retrieval. *Cerebral Cortex*, 14(4), 364–375. <http://www.ncbi.nlm.nih.gov/pubmed/15028641>
- Calahorra, F., & Ruiz-Rubio, M. (2011). *Caenorhabditis elegans* as an experimental tool for the study of complex neurological diseases: Parkinson's disease, Alzheimer's disease and autism spectrum disorder. In *Invertebrate Neuroscience* (Vol. 11, Issue 2, pp. 73–83). Invert Neurosci. <https://doi.org/10.1007/s10158-011-0126-1>
- Caliva, F., Iriondo, C., Martinez, A. M., Majumdar, S., & Pedoia, V. (2019). *Distance Map Loss Penalty Term for Semantic Segmentation*. <http://arxiv.org/abs/1908.03679>
- Callaway, E. M. (2008). Transneuronal circuit tracing with neurotropic viruses. In *Current Opinion in Neurobiology* (Vol. 18, Issue 6, pp. 617–623). Elsevier Current Trends. <https://doi.org/10.1016/j.conb.2009.03.007>
- Callaway, E. M., & Luo, L. (2015). Monosynaptic circuit tracing with glycoprotein-deleted rabies viruses. *Journal of Neuroscience*, 35(24), 8979–8985. <https://doi.org/10.1523/JNEUROSCI.0409-15.2015>
- Card, J. P., Rinaman, L., Schwaber, J. S., Miselis, R. R., Whealy, M. E., Robbins, A. K., & Enquist, L. W. (1990). Neurotropic properties of pseudorabies virus: Uptake and transneuronal passage in the rat central nervous system. *Journal of Neuroscience*, 10(6), 1974–1994. <https://doi.org/10.1523/jneurosci.10-06-01974.1990>
- Cattell, R. B. (1987). *Intelligence : Its structure, growth, and action*. North-Holland.
- Cepeda-Prado, E., Popp, S., Khan, U., Stefanov, D., Rodríguez, J., Menalled, L. B., Dow-Edwards, D., Small, S. A., & Moreno, H. (2012). R6/2 huntington's disease mice develop early and progressive abnormal brain metabolism and seizures. *Journal of Neuroscience*, 32(19), 6456–6467. <https://doi.org/10.1523/JNEUROSCI.0388-12.2012>
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., & Prasher, D. C. (1994). Green fluorescent

- protein as a marker for gene expression. *Science*, 263(5148), 802–805.  
<https://doi.org/10.1126/science.8303295>
- Chávez, M. N., Aedo, G., Fierro, F. A., Allende, M. L., & Egaña, J. T. (2016). Zebrafish as an emerging model organism to study angiogenesis in development and regeneration. In *Frontiers in Physiology* (Vol. 7, Issue MAR). Frontiers Media S.A.  
<https://doi.org/10.3389/fphys.2016.00056>
- Chollet, F., & others. (2015). *Keras*. GitHub.
- Chon, U., Vanselow, D. J., Cheng, K. C., & Kim, Y. (2019). Enhanced and unified anatomical labeling for a common mouse brain atlas. *Nature Communications*, 10(1), 1–12.  
<https://doi.org/10.1038/s41467-019-13057-w>
- Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., D’Anastasi, M., Sommer, W. H., Ahmadi, S. A., & Menze, B. H. (2016). Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9901 LNCS, 415–423. [https://doi.org/10.1007/978-3-319-46723-8\\_48](https://doi.org/10.1007/978-3-319-46723-8_48)
- Chung, K., & Deisseroth, K. (2013). CLARITY for mapping the nervous system. *Nature Methods*, 10(6), 508–513. <https://doi.org/10.1038/nmeth.2481>
- Chung, K., Wallace, J., Kim, S.-Y., Kalyanasundaram, S., Andalman, A. S., Davidson, T. J., Mirzabekov, J. J., Zalocusky, K. A., Mattis, J., Denisin, A. K., Pak, S., Bernstein, H., Ramakrishnan, C., Grosenick, L., Gradinaru, V., & Deisseroth, K. (2013). Structural and molecular interrogation of intact biological systems. *Nature*, 497(7449), 332–337.  
<https://doi.org/10.1038/nature12107>
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9901 LNCS, 424–432. <http://arxiv.org/abs/1606.06650>
- Ciftcioglu, U. M., Suresh, V., Ding, K. R., Sommer, F. T., & Hirsch, J. A. (2020). Visual information processing in the ventral division of the mouse lateral geniculate nucleus of the thalamus. *The Journal of Neuroscience*, JN-RM-2602-19.  
<https://doi.org/10.1523/JNEUROSCI.2602-19.2020>



- Cireřan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column Deep Neural Networks for Image Classification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3642–3649. <http://arxiv.org/abs/1202.2745>
- Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2015). Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*. <http://arxiv.org/abs/1511.07289>
- Courtiol, E., & Wilson, D. A. (2015). The olfactory thalamus: Unanswered questions about the role of the mediodorsal thalamic nucleus in olfaction. *Frontiers in Neural Circuits*, 9(September). <https://doi.org/10.3389/fncir.2015.00049>
- Cruz-Martín, A., El-Danaf, R. N., Osakada, F., Sriram, B., Dhande, O. S., Nguyen, P. L., Callaway, E. M., Ghosh, A., & Huberman, A. D. (2014). A dedicated circuit links direction-selective retinal ganglion cells to the primary visual cortex. *Nature*, 507(7492), 358–361. <https://doi.org/10.1038/nature12989>
- De Giorgio, F., Maduro, C., Fisher, E. M. C., & Acevedo-Arozena, A. (2019). Transgenic and physiological mouse models give insights into different aspects of amyotrophic lateral sclerosis. In *DMM Disease Models and Mechanisms* (Vol. 12, Issue 1). Company of Biologists Ltd. <https://doi.org/10.1242/dmm.037424>
- De Groot, J. C., De Leeuw, F.-E., Oudkerk, M., Van Gijn, J., Hofman, A., Jolles, J., & Breteler, M. M. B. (2002). Periventricular cerebral white matter lesions predict rate of cognitive decline. *Annals of Neurology*, 52(3), 335–341. <https://doi.org/10.1002/ana.10294>
- De Groot, J. C., De Leeuw, F. E., Oudkerk, M., van Gijn, J., Hofman, A., Jolles, J., & Breteler, M. M. (2000). Cerebral white matter lesions and cognitive function: the Rotterdam Scan Study. *Annals of Neurology*, 47(2), 145–151. <http://www.ncbi.nlm.nih.gov/pubmed/10665484>
- De Leeuw, F. E., De Groot, J. C., Achten, E., Oudkerk, M., Ramos, L. M. P., Heijboer, R., Hofman, A., Jolles, J., Van Gijn, J., & Breteler, M. M. B. (2001). Prevalence of cerebral white matter lesions in elderly people: A population based magnetic resonance imaging study. The Rotterdam Scan Study. *Journal of Neurology Neurosurgery and Psychiatry*, 70(1), 9–14. <https://doi.org/10.1136/jnnp.70.1.9>
- Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., Penke, L., Rafnsson, S. B., & Starr, J. M. (2009). Age-associated cognitive decline. *British Medical*

- Bulletin*, 92(1), 135–152. <https://doi.org/10.1093/bmb/ldp033>
- Deary, I. J., Pattie, A., & Starr, J. M. (2013). The Stability of Intelligence From Age 11 to Age 90 Years: The Lothian Birth Cohort of 1921. *Psychological Science*, 24(12), 2361–2368. <https://doi.org/10.1177/0956797613486487>
- Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish mental survey. In *Intelligence* (Vol. 28, pp. 49–55).
- Deary, I. J., Whiteman, M., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947 . In *Journal of Personality and Social Psychology* (Vol. 86, Issue 1, pp. 130–147). <https://doi.org/10.1037/0022-3514.86.1.130>
- Delogu, A., Sellers, K., Zagoraiou, L., Bocianowska-Zbrog, A., Mandal, S., Guimera, J., Rubenstein, J. L. R., Sugden, D., Jessell, T., & Lumsden, A. (2012). Subcortical Visual Shell Nuclei Targeted by ipRGCs Develop from a Sox14+-GABAergic Progenitor and Require Sox14 to Regulate Daily Activity Rhythms. *Neuron*, 75(4), 648–662. <https://doi.org/10.1016/j.neuron.2012.06.013>
- Dembrow, N., & Johnston, D. (2014). Subcircuit-specific neuromodulation in the prefrontal cortex. In *Frontiers in Neural Circuits* (Vol. 8, Issue JUNE). Frontiers Research Foundation. <https://doi.org/10.3389/fncir.2014.00054>
- Denk, W., Strickler, J. H., & Webb, W. W. (1990). Two-photon laser scanning fluorescence microscopy. *Science (New York, N.Y.)*, 248(4951), 73–76. <https://doi.org/10.1126/science.2321027>
- Ding, S.-L., Royall, J. J., Sunkin, S. M., Ng, L., Facer, B. A. C., Lesnar, P., Guillozet-Bongaarts, A., McMurray, B., Szafer, A., Dolbeare, T. A., Stevens, A., Tirrell, L., Benner, T., Caldejon, S., Dalley, R. A., Dee, N., Lau, C., Nyhus, J., Reding, M., ... Lein, E. S. (2016). Comprehensive cellular-resolution atlas of the adult human brain. *Journal of Comparative Neurology*, 524(16), 3127–3481. <https://doi.org/10.1002/cne.24080>
- Dmitrieff, S., & Nédélec, F. (2017). ConfocalGN: A minimalistic confocal image generator. *SoftwareX*, 6, 243–247. <https://doi.org/10.1016/J.SOFTX.2017.09.002>
- Donehower, L. A., Harvey, M., Slagle, B. L., McArthur, M. J., Montgomery, C. A., Butel, J. S., & Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature*, 356(6366), 215–221.

<https://doi.org/10.1038/356215a0>

Dong, H. W. (2008). The Allen reference atlas: A digital color brain atlas of the C57Bl/6J male mouse. In *The Allen reference atlas: A digital color brain atlas of the C57Bl/6J male mouse*. John Wiley & Sons Inc.

Drobizhev, M., Makarov, N. S., Tillo, S. E., Hughes, T. E., & Rebane, A. (2011). Two-photon absorption properties of fluorescent proteins. *Nature Methods*, 8(5), 393–399.

<https://doi.org/10.1038/nmeth.1596>

Durr, N. J., Weisspfennig, C. T., Holfeld, B. A., & Ben-Yakar, A. (2011). Maximum imaging depth of two-photon autofluorescence microscopy in epithelial tissues. *Journal of Biomedical Optics*, 16(2), 026008. <https://doi.org/10.1117/1.3548646>

Ellis, E. M., Gauvain, G., Sivyer, B., & Murphy, G. J. (2016). Shared and distinct retinal input to the mouse superior colliculus and dorsal lateral geniculate nucleus. *Journal of Neurophysiology*, 116(2), 602–610. <https://doi.org/10.1152/jn.00227.2016>

Emery, L., Heaven, T. J., Paxton, J. L., & Braver, T. S. (2008). Age-related changes in neural activity during performance matched working memory manipulation. *NeuroImage*, 42(4), 1577–1586. <https://doi.org/10.1016/j.neuroimage.2008.06.021>

Ertürk, A., Becker, K., Jährling, N., Mauch, C. P., Hojer, C. D., Egen, J. G., Hellal, F., Bradke, F., Sheng, M., & Dodt, H.-U. (2012). Three-dimensional imaging of solvent-cleared organs using 3DISCO. *Nature Protocols*, 7(11), 1983–1995.

<https://doi.org/10.1038/nprot.2012.119>

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>

Evangelio, M., García-Amado, M., & Clascá, F. (2018). Thalamocortical Projection Neuron and Interneuron Numbers in the Visual Thalamic Nuclei of the Adult C57BL/6 Mouse. *Frontiers in Neuroanatomy*, 12, 27. <https://doi.org/10.3389/fnana.2018.00027>

Farnum, A., & Pelled, G. (2020). New Vision for Visual Prostheses . In *Frontiers in Neuroscience* (Vol. 14, p. 36).

<https://www.frontiersin.org/article/10.3389/fnins.2020.00036>

Fonseca, M. S., Bergomi, M. G., Mainen, Z. F., & Shemesh, N. (2020). Functional MRI of large scale activity in behaving mice. *BioRxiv*, 2020.04.16.044941.

<https://doi.org/10.1101/2020.04.16.044941>

- Franklin, K., & Paxinos, G. (2008). *The Mouse Brain in Stereotaxic Coordinates. The Coronal Plates and Diagrams* (K. Franklin & G. Paxinos (eds.)). Academic Press.
- Frasconi, P., Silvestri, L., Soda, P., Cortini, R., Pavone, F. S., & Iannello, G. (2014). Large-scale automated identification of mouse brain cells in confocal light sheet microscopy images. *Bioinformatics*, 30(17), i587–i593.  
<https://doi.org/10.1093/bioinformatics/btu469>
- Frisoni, G. B., Testa, C., Zorzan, A., Sabattoli, F., Beltramello, A., Soininen, H., & Laakso, M. P. (2002). Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry. *Journal of Neurology, Neurosurgery, and Psychiatry*, 73(6), 657–664.  
<http://www.ncbi.nlm.nih.gov/pubmed/12438466>
- Fu, Z. F., & Jackson, A. C. (2005). Neuronal dysfunction and death in rabies virus infection. In *Journal of NeuroVirology* (Vol. 11, Issue 1, pp. 101–106). J Neurovirol.  
<https://doi.org/10.1080/13550280590900445>
- Ge, Y., Grossman, R. I., Babb, J. S., Rabin, M. L., Mannon, L. J., & Kolson, D. L. (2002). Age-related total gray matter and white matter changes in normal adult brain. Part I: volumetric MR imaging analysis. *AJNR. American Journal of Neuroradiology*, 23(8), 1327–1333. <http://www.ncbi.nlm.nih.gov/pubmed/12223373>
- Genzel, L., Schut, E., Schröder, T., Eichler, R., Khamassi, M., Gomez, A., Navarro Lobato, I., & Battaglia, F. (2019). The object space task shows cumulative memory expression in both mice and rats. *PLOS Biology*, 17(6), e3000322.  
<https://doi.org/10.1371/journal.pbio.3000322>
- Girkin, J. M., & Carvalho, M. T. (2018). The light-sheet microscopy revolution. *Journal of Optics*, 20(5), 53002. <https://doi.org/10.1088/2040-8986/aab58a>
- Godement, P., Salaün, J., & Imbert, M. (1984). Prenatal and postnatal development of retinogeniculate and retinocollicular projections in the mouse. *Journal of Comparative Neurology*, 230(4), 552–575. <https://doi.org/10.1002/cne.902300406>
- Golding, B., Pouchelon, G., Bellone, C., Murthy, S., Di Nardo, A. A., Govindan, S., Ogawa, M., Shimogori, T., Lüscher, C., Dayer, A., & Jabaudon, D. (2014). Retinal input directs the recruitment of inhibitory interneurons into thalamic visual circuits. *Neuron*, 81(5), 1057–1069. <https://doi.org/10.1016/j.neuron.2014.01.032>
- Gong, H., Xu, D., Yuan, J., Li, X., Guo, C., Peng, J., Li, Y., Schwarz, L. A., Li, A., Hu, B., Xiong, B., Sun, Q., Zhang, Y., Liu, J., Zhong, Q., Xu, T., Zeng, S., & Luo, Q. (2016). High-throughput

- dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nature Communications*, 7(1), 1–12.  
<https://doi.org/10.1038/ncomms12142>
- Gong, H., Zeng, S., Yan, C., Lv, X., Yang, Z., Xu, T., Feng, Z., Ding, W., Qi, X., Li, A., Wu, J., & Luo, Q. (2013). Continuously tracing brain-wide long-distance axonal projections in mice at a one-micron voxel resolution. *NeuroImage*, 74, 87–98.  
<https://doi.org/10.1016/j.neuroimage.2013.02.005>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*.
- Göppert-Mayer, M. (2009). Elementary processes with two quantum transitions. *Annalen Der Physik*, 18(7–8), 466–479. <https://doi.org/10.1002/andp.200910358>
- Goyal, N., Siddiqui, S., Chatterjee, U., Kumar, D., & Siddiqui, A. (2008). Neuropsychology of prefrontal cortex. *Indian Journal of Psychiatry*, 50(3), 202.  
<https://doi.org/10.4103/0019-5545.43634>
- Graham, L. C., Grabowska, W. A., Chun, Y., Risacher, S. L., Philip, V. M., Saykin, A. J., Sukoff Rizzo, S. J., & Howell, G. R. (2019). Exercise prevents obesity-induced cognitive decline and white matter damage in mice. *Neurobiology of Aging*, 80, 154–172.  
<https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2019.03.018>
- Grewe, B. F., Langer, D., Kasper, H., Kampa, B. M., & Helmchen, F. (2010). High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nature Methods*, 7(5), 399–405. <https://doi.org/10.1038/nmeth.1453>
- Guido, W. (2018). Development, form, and function of the mouse visual thalamus. In *Journal of Neurophysiology* (Vol. 120, Issue 1, pp. 211–225). American Physiological Society.  
<https://doi.org/10.1152/jn.00651.2017>
- Gunning-Dixon, F. M., Brickman, A. M., Cheng, J. C., & Alexopoulos, G. S. (2009). Aging of cerebral white matter: A review of MRI findings. In *International Journal of Geriatric Psychiatry* (Vol. 24, Issue 2, pp. 109–117). NIH Public Access.  
<https://doi.org/10.1002/gps.2087>
- Gunning-Dixon, F. M., & Raz, N. (2000). The cognitive correlates of white matter abnormalities in normal aging: A quantitative review. *Neuropsychology*, 14(2), 224–232. <https://doi.org/10.1037/0894-4105.14.2.224>
- Hackett, T. A., Barkat, T. R., O'Brien, B. M. J., Hensch, T. K., & Polley, D. B. (2011). Linking

- topography to tonotopy in the mouse auditory thalamocortical circuit. *Journal of Neuroscience*, 31(8), 2983–2995. <https://doi.org/10.1523/JNEUROSCI.5333-10.2011>
- Hall, D. H., Hartweg, E., & Nguyen, K. C. Q. (2012). Modern Electron Microscopy Methods for *C. elegans*. In *Methods in Cell Biology* (Vol. 107, pp. 93–149). Academic Press Inc. <https://doi.org/10.1016/B978-0-12-394620-1.00004-7>
- Hammelrath, L., Škokić, S., Khmelinskii, A., Hess, A., van der Knaap, N., Staring, M., Lelieveldt, B. P. F., Wiedermann, D., & Hoehn, M. (2016). Morphological maturation of the mouse brain: An in vivo MRI and histology investigation. *NeuroImage*, 125, 144–152. <https://doi.org/10.1016/j.neuroimage.2015.10.009>
- Han, J., & Moraga, C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 930, 195–201. [https://doi.org/10.1007/3-540-59497-3\\_175](https://doi.org/10.1007/3-540-59497-3_175)
- Harrington, A. J., Hamamichi, S., Caldwell, G. A., & Caldwell, K. A. (2010). *C. elegans* as a model organism to investigate molecular pathways involved with Parkinson's disease. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, 239(5), 1282–1295. <https://doi.org/10.1002/dvdy.22231>
- Harrington, M. E. (1997). The ventral lateral geniculate nucleus and the intergeniculate leaflet: Interrelated structures in the visual and circadian systems. *Neuroscience and Biobehavioral Reviews*, 21(5), 705–727. [https://doi.org/10.1016/S0149-7634\(96\)00019-X](https://doi.org/10.1016/S0149-7634(96)00019-X)
- Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2), 100–107. <https://doi.org/10.1109/TSSC.1968.300136>
- Helmchen, F., & Denk, W. (2005). Deep tissue two-photon microscopy. *Nature Methods*, 2(12), 932–940. <https://doi.org/10.1038/nmeth818>
- Herculano-Houzel, S., Watson, C., & Paxinos, G. (2013). Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones. *Frontiers in Neuroanatomy*, 7(OCT), 35. <https://doi.org/10.3389/fnana.2013.00035>
- Hildebrand, D. G. C., Cicconet, M., Torres, R. M., Choi, W., Quan, T. M., Moon, J., Wetzel, A. W., Scott Champion, A., Graham, B. J., Randlett, O., Plummer, G. S., Portugues, R., Bianco, I. H., Saalfeld, S., Baden, A. D., Lillanay, K., Burns, R., Vogelstein, J. T., Schier, A.

- F., ... Engert, F. (2017). Whole-brain serial-section electron microscopy in larval zebrafish. *Nature*, 545(7654), 345–349. <https://doi.org/10.1038/nature22356>
- Hodges, H. (1996). Maze procedures: The radial-arm and water maze compared. *Cognitive Brain Research*, 3(3–4), 167–181. [https://doi.org/10.1016/0926-6410\(96\)00004-3](https://doi.org/10.1016/0926-6410(96)00004-3)
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4), 500–544. <https://doi.org/10.1113/jphysiol.1952.sp004764>
- Hoff, E. C. (1932). Central nerve terminals in the mamalian spinal cord and their examination by experimental degeneration. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 111(770), 175–188. <https://doi.org/10.1098/rspb.1932.0049>
- Höller-Wallscheid, M. S., Thier, P., Pomper, J. K., & Lindner, A. (2017). Bilateral recruitment of prefrontal cortex in working memory is associated with task demand but not with age. *Proceedings of the National Academy of Sciences of the United States of America*, 114(5), E830–E839. <https://doi.org/10.1073/pnas.1601983114>
- Home Office. (2019). *Annual Statistics of Scientific Procedures on Living Animals Great Britain 2018*.
- Horton, N. G., Wang, K., Kobat, D., Clark, C. G., Wise, F. W., Schaffer, C. B., & Xu, C. (2013). In vivo three-photon microscopy of subcortical structures within an intact mouse brain. *Nature Photonics*, 7(3), 205–209. <https://doi.org/10.1038/nphoton.2012.336>
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G. J., White, S., ... Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446), 498–503. <https://doi.org/10.1038/nature12111>
- Huisken, J., & Stainier, D. Y. R. (2007). Even fluorescence excitation by multidirectional selective plane illumination microscopy (mSPIM). *Optics Letters*, 32(17), 2608. <https://doi.org/10.1364/ol.32.002608>
- Hüsemann, Y., Geigl, J. B., Schubert, F., Musiani, P., Meyer, M., Burghart, E., Forni, G., Eils, R., Fehm, T., Riethmüller, G., & Klein, C. A. (2008). Systemic Spread Is an Early Step in Breast Cancer. *Cancer Cell*, 13(1), 58–68. <https://doi.org/10.1016/j.ccr.2007.12.003>
- Immerkær, J. (1996). Fast noise variance estimation. *Computer Vision and Image*

- Understanding*, 64(2), 300–302. <https://doi.org/10.1006/cviu.1996.0060>
- Iqbal, A., Khan, R., & Karayannis, T. (2019). Developing a brain atlas through deep learning. *Nature Machine Intelligence*, 1(6), 277–287. <https://doi.org/10.1038/s42256-019-0058-8>
- Isobe, K., Toda, K., Song, Q., Kannari, F., Kawano, H., Miyawaki, A., & Midorikawa, K. (2017). Temporal focusing microscopy combined with three-dimensional structured illumination. *Japanese Journal of Applied Physics*, 56(5), 052501. <https://doi.org/10.7567/JJAP.56.052501>
- Jager, P., Moore, G., Calpin, P., Durmishi, X., Kita, Y., Salgarella, I., Wang, Y., Schultz, S. R., Brickley, S., Shimogori, T., & Delogu, A. (2020). Dual midbrain and forebrain origins of thalamic inhibitory interneurons. *BioRxiv*.
- Jager, P., Ye, Z., Yu, X., Zagoraïou, L., Prekop, H. T., Partanen, J., Jessell, T. M., Wisden, W., Brickley, S. G., & Delogu, A. (2016). Tectal-derived interneurons contribute to phasic and tonic inhibition in the visual thalamus. *Nature Communications*, 7, 1–14. <https://doi.org/10.1038/ncomms13579>
- Jaubert-Miazza, L., Green, E., Lo, F. S., Bui, K., Mills, J., & Guido, W. (2005). Structural and functional composition of the developing retinogeniculate pathway in the mouse. *Visual Neuroscience*, 22(5), 661–676. <https://doi.org/10.1017/S0952523805225154>
- Jawhar, S., Trawicka, A., Jenneckens, C., Bayer, T. A., & Wirths, O. (2012). Motor deficits, neuron loss, and reduced anxiety coinciding with axonal degeneration and intraneuronal A $\beta$  aggregation in the 5XFAD mouse model of Alzheimer’s disease. *Neurobiology of Aging*, 33(1), 196.e29-196.e40. <https://doi.org/10.1016/j.neurobiolaging.2010.05.027>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Jernigan, T. L., Archibald, S. L., Fennema-Notestine, C., Gamst, A. C., Stout, J. C., Bonner, J., & Hesselink, J. R. (2001). Effects of age on tissues and regions of the cerebrum and cerebellum. *Neurobiology of Aging*, 22(4), 581–594. <http://www.ncbi.nlm.nih.gov/pubmed/11445259>
- Jinno, S., & Kosaka, T. (2010). Stereological estimation of numerical densities of glutamatergic principal neurons in the mouse hippocampus. *Hippocampus*, 20(7), 829–840. <https://doi.org/10.1002/hipo.20685>



- Johnson, G. A., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S., & Nissanov, J. (2010). Waxholm Space: An image-based reference for coordinating mouse brain research. *NeuroImage*, 53(2), 365–372. <https://doi.org/10.1016/j.neuroimage.2010.06.067>
- Jonkman, J., & Brown, C. M. (2015). Any way you slice it—A comparison of confocal microscopy techniques. *Journal of Biomolecular Techniques*, 26(2), 54–65. <https://doi.org/10.7171/jbt.15-2602-003>
- Jurgens, C. W. D., Bell, K. A., McQuiston, A. R., & Guido, W. (2012). Optogenetic Stimulation of the Corticothalamic Pathway Affects Relay Cells and GABAergic Neurons Differently in the Mouse Visual Thalamus. *PLoS ONE*, 7(9). <https://doi.org/10.1371/journal.pone.0045717>
- Karnovsky, M., & Roots, L. (1964). A “direct-coloring” method for cholinesterases. *Journal of Histochemistry and Cytochemistry. The Journal of Histochemistry and Cytochemistry : Official Journal of the Histochemistry Society*, 12, 219–221. <http://www.ncbi.nlm.nih.gov/pubmed/14187330>
- Kawaguchi, Y., & Kubota, Y. (1998). Neurochemical features and synaptic connections of large physiologically-identified GABAergic cells in the rat frontal cortex. *Neuroscience*, 85(3), 677–701. [https://doi.org/10.1016/S0306-4522\(97\)00685-4](https://doi.org/10.1016/S0306-4522(97)00685-4)
- Keller, D., Erö, C., & Markram, H. (2018). Cell densities in the mouse brain: A systematic review. *Frontiers in Neuroanatomy*, 12(October). <https://doi.org/10.3389/fnana.2018.00083>
- Keller, P. J., & Dodt, H. U. (2012). Light sheet microscopy of living or cleared specimens. In *Current Opinion in Neurobiology* (Vol. 22, Issue 1, pp. 138–143). <https://doi.org/10.1016/j.conb.2011.08.003>
- 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](https://doi.org/10.1016/S0165-0270(00)00296-X)
- Kerschensteiner, D., & Guido, W. (2017). Organization of the dorsal lateral geniculate nucleus in the mouse. *Visual Neuroscience*, 34, E008–E008. <https://doi.org/10.1017/S0952523817000062>
- Kim, J., Zhao, T., Petralia, R. S., Yu, Y., Peng, H., Myers, E., & Magee, J. C. (2011). mGRASP enables mapping mammalian synaptic connectivity with light microscopy. *Nature Methods*, 9(1), 96–102. <https://doi.org/10.1038/nmeth.1784>

- Kim, Y., Venkataraju, K. U., Pradhan, K., Mende, C., Taranda, J., Turaga, S. C., Arganda-Carreras, I., Ng, L., Hawrylycz, M. J., Rockland, K. S., Seung, H. S., & Osten, P. (2015). Mapping social behavior-induced brain activation at cellular resolution in the mouse. *Cell Reports*, 10(2), 292–305. <https://doi.org/10.1016/j.celrep.2014.12.014>
- King, D. L., & Arendash, G. W. (2002). Behavioral characterization of the Tg2576 transgenic model of Alzheimer's disease through 19 months. *Physiology and Behavior*, 75(5), 627–642. [https://doi.org/10.1016/S0031-9384\(02\)00639-X](https://doi.org/10.1016/S0031-9384(02)00639-X)
- Kinugawa, K., Schumm, S., Pollina, M., Depre, M., Jungbluth, C., Doulazmi, M., Sebban, C., Zlomuzica, A., Pietrowsky, R., Pause, B., Mariani, J., & Dere, E. (2013). Aging-related episodic memory decline: Are emotions the key? *Frontiers in Behavioral Neuroscience*, 7(JANUARY 2013). <https://doi.org/10.3389/fnbeh.2013.00002>
- Klein, S., Staring, M., Murphy, K., Viergever, M. M. A., & Pluim, J. (2009). 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>
- Kner, P., Sedat, J. W., Agard, D. A., & Kam, Z. (2010). High-resolution wide-field microscopy with adaptive optics for spherical aberration correction and motionless focusing. *Journal of Microscopy*, 237(2), 136–147. <https://doi.org/10.1111/j.1365-2818.2009.03315.x>
- Kobat, D., Horton, N. G., & Xu, C. (2011). In vivo two-photon microscopy to 1.6-mm depth in mouse cortex. *Journal of Biomedical Optics*, 16(10), 106014. <https://doi.org/10.1117/1.3646209>
- Kobayashi, H., Ogawa, M., Alford, R., Choyke, P. L., & Urano, Y. (2010). New strategies for fluorescent probe design in medical diagnostic imaging. *Chemical Reviews*, 110(5), 2620–2640. <https://doi.org/10.1021/cr900263j>
- Kourkoutis, L. F., Plitzko, J. M., & Baumeister, W. (2012). Electron Microscopy of Biological Materials at the Nanometer Scale. *Annual Review of Materials Research*, 42(1), 33–58. <https://doi.org/10.1146/annurev-matsci-070511-155004>
- Krahe, T. E., El-Danaf, R. N., Dilger, E. K., Henderson, S. C., & Guido, W. (2011). Morphologically distinct classes of relay cells exhibit regional preferences in the dorsal lateral geniculate nucleus of the mouse. *Journal of Neuroscience*, 31(48), 17437–17448. <https://doi.org/10.1523/JNEUROSCI.4370-11.2011>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep

- convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.  
<https://doi.org/10.1145/3065386>
- Kuan, L., Li, Y., Lau, C., Feng, D., Bernard, A., Sunkin, S. M., Zeng, H., Dang, C., Hawrylycz, M., & Ng, L. (2015). Neuroinformatics of the allen mouse brain connectivity atlas. *Methods*, 73, 4–17. <https://doi.org/10.1016/j.ymeth.2014.12.013>
- Larsen, D. L. D., & Callaway, E. M. (2006). Development of layer-specific axonal arborizations in mouse primary somatosensory cortex. *Journal of Comparative Neurology*, 494(3), 398–414. <https://doi.org/10.1002/cne.20754>
- Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394), 190–191.  
<https://doi.org/10.1038/242190a0>
- Lawrence, P. M., & Studholme, K. M. (2014). Retinofugal projections in the mouse. *Journal of Comparative Neurology*, 522(16), 3733–3753. <https://doi.org/10.1002/cne.23635>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2323.  
<https://doi.org/10.1109/5.726791>
- Leist, M., Datunashvilli, M., Kanyshkova, T., Zobeiri, M., Aissaoui, A., Cerina, M., Romanelli, M. N., Pape, H. C., & Budde, T. (2016). Two types of interneurons in the mouse lateral geniculate nucleus are characterized by different h-current density. *Scientific Reports*, 6. <https://doi.org/10.1038/srep24904>
- Lenehan, M. E., Summers, M. J., Saunders, N. L., Summers, J. J., & Vickers, J. C. (2015). Relationship between education and age-related cognitive decline: A review of recent research. In *Psychogeriatrics* (Vol. 15, Issue 2, pp. 154–162). Blackwell Publishing Ltd.  
<https://doi.org/10.1111/psyg.12083>
- Leonard, D. N., Chandler, G. W., & Seraphin, S. (2012). Scanning Electron Microscopy. In *Characterization of Materials* (pp. 1–16). John Wiley & Sons, Inc.  
<https://doi.org/10.1002/0471266965.com081.pub2>
- LeVay, S., & Ferster, D. (1979). Proportion of interneurons in the cat's lateral geniculate nucleus. *Brain Research*, 164(1–2), 304–308. [https://doi.org/10.1016/0006-8993\(79\)90026-X](https://doi.org/10.1016/0006-8993(79)90026-X)
- Lewiner, T., Lopes, H., Vieira, A. W., & Tavares, G. (2003). Efficient Implementation of Marching Cubes' Cases with Topological Guarantees. *Journal of Graphics Tools*, 8(2), 1–

15. <https://doi.org/10.1080/10867651.2003.10487582>
- Lezak, K. R., Missig, G., & Carlezon, W. A. (2017). Behavioral methods to study anxiety in rodents. *Dialogues in Clinical Neuroscience*, 19(2), 181–191.
- Li, A., Gong, H., Zhang, B., Wang, Q., Yan, C., Wu, J., Liu, Q., Zeng, S., & Luo, Q. (2010). Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain. *Science*, 330(6009), 1404–1408. <https://doi.org/10.1126/science.1191776>
- Lin, G., Adiga, U., Olson, K., Guzowski, J. F., Barnes, C. A., & Roysam, B. (2003). A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry*, 56A(1), 23–36. <https://doi.org/10.1002/cyto.a.10079>
- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2020). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 318–327. <https://doi.org/10.1109/TPAMI.2018.2858826>
- Liu, H., Wang, L., Geng, Z., Zhu, Q., Song, Z., Chang, R., & Lv, H. (2016). A voxel-based morphometric study of age- and sex-related changes in white matter volume in the normal aging brain. *Neuropsychiatric Disease and Treatment*, 12, 453–465. <https://doi.org/10.2147/NDT.S90674>
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150–157. <https://doi.org/10.1038/35084005>
- Longair, M. H., Baker, D. A., & Armstrong, J. D. (2011). Simple Neurite Tracer: open source software for reconstruction, visualization and analysis of neuronal processes. *Bioinformatics*, 27(17), 2453–2454. <https://doi.org/10.1093/bioinformatics/btr390>
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Lowekamp, B. C., Chen, D. T., Ibáñez, L., & Blezek, D. (2013). The design of simpleITK. *Frontiers in Neuroinformatics*, 7(DEC). <https://doi.org/10.3389/fninf.2013.00045>
- Lukaschek, K., Vanajan, A., Johar, H., Weiland, N., & Ladwig, K. H. (2017). “In the mood for ageing”: Determinants of subjective well-being in older men and women of the population-based KORA-Age study. *BMC Geriatrics*, 17(1). <https://doi.org/10.1186/s12877-017-0513-5>

- Madisen, L., Garner, A. R., Shimaoka, D., Chuong, A. S., Klapoetke, N. C., Li, L., Van Der Bourg, A., Niino, Y., Egolf, L., Monetti, C., Gu, H., Mills, M., Cheng, A., Tasic, B., Nguyen, T. N., Sunkin, S. M., Benucci, A., Nagy, A., Miyawaki, A., ... Zeng, H. (2015). Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron*, 85(5), 942–958.  
<https://doi.org/10.1016/j.neuron.2015.02.022>
- Maeda, Y., Kurokawa, K., Ito, Y., Nakano, A., & Wada, S. (2018). Wide field video-rate two-photon imaging by using spinning disk beam scanner. In A. Periasamy, P. T. So, X. S. Xie, & K. König (Eds.), *Multiphoton Microscopy in the Biomedical Sciences XVIII* (Vol. 10498, p. 88). SPIE. <https://doi.org/10.1117/12.2288969>
- Maheswaran, S., Barjat, H., Rueckert, D., Bate, S. T., Howlett, D. R., Tilling, L., Smart, S. C., Pohlmann, A., Richardson, J. C., Hartkens, T., Hill, D. L. G., Upton, N., Hajnal, J. V., & James, M. F. (2009). Longitudinal regional brain volume changes quantified in normal aging and Alzheimer's APP × PS1 mice using MRI. *Brain Research*, 1270, 19–32.  
<https://doi.org/10.1016/j.brainres.2009.02.045>
- Martin, P. R. (1986). The projection of different retinal ganglion cell classes to the dorsal lateral geniculate nucleus in the hooded rat. *Experimental Brain Research*, 62(1), 77–88. <https://doi.org/10.1007/BF00237404>
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9), 1281–1289. <https://doi.org/10.1038/s41593-018-0209-y>
- Mccall, R. B. (1977). Childhood iq's as predictors of adult educational and occupational status. *Science*, 197(4302), 482–483. <https://doi.org/10.1126/science.197.4302.482>
- McKinnon, K. M. (2018). Flow cytometry: An overview. *Current Protocols in Immunology*, 2018, 5.1.1-5.1.11. <https://doi.org/10.1002/cpim.40>
- Meier-Ruge, W., Ulrich, J., Brühlmann, M., & Meier, E. (1992). Age-related white matter atrophy in the human brain. *Annals of the New York Academy of Sciences*, 673, 260–269. <http://www.ncbi.nlm.nih.gov/pubmed/1485724>
- Merkow, J., Kriegman, D., Marsden, A., & Tu, Z. (2016). Dense Volume-to-Volume Vascular Boundary Detection. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9902 LNCS, 371–

379. <http://arxiv.org/abs/1605.08401>
- Min, Z., Jiguo, Z., & Xusheng, X. (2012). Panorama Stitching Based on SIFT Algorithm and Levenberg-Marquardt Optimization. *Physics Procedia*, 33, 811–818.  
<https://doi.org/10.1016/j.phpro.2012.05.139>
- Minsky, M. (1988). Memoir on inventing the confocal scanning microscope. *Scanning*, 10(4), 128–138. <https://doi.org/10.1002/sca.4950100403>
- Miranda, R., Sébrié, C., Degrouard, J., Gillet, B., Jaillard, D., Laroche, S., & Vaillend, C. (2009). Reorganization of inhibitory synapses and increased PSD length of perforated excitatory synapses in hippocampal area CA1 of dystrophin-deficient mdx mice. *Cerebral Cortex (New York, N.Y. : 1991)*, 19(4), 876–888.  
<https://doi.org/10.1093/cercor/bhn135>
- Mirzoyan, Z., Sollazzo, M., Allocca, M., Valenza, A. M., Grifoni, D., & Bellosta, P. (2019). *Drosophila melanogaster*: A model organism to study cancer. In *Frontiers in Genetics* (Vol. 10, p. 51). Frontiers Media S.A. <https://doi.org/10.3389/fgene.2019.00051>
- Modat, M., Cash, D. M., Daga, P., Winston, G. P., Duncan, J. S., & Ourselin, S. (2014). Global image registration using a symmetric block-matching approach. *Journal of Medical Imaging*, 1(2), 024003. <https://doi.org/10.1117/1.jmi.1.2.024003>
- Monavarfeshani, A., Sabbagh, U., & Fox, M. A. (2017). Not a one-trick pony: Diverse connectivity and functions of the rodent lateral geniculate complex. *Visual Neuroscience*, 34(January), E012. <https://doi.org/10.1017/S0952523817000098>
- Monici, M. (2005). Cell and tissue autofluorescence research and diagnostic applications. In *Biotechnology Annual Review* (Vol. 11, Issue SUPPL., pp. 227–256).  
[https://doi.org/10.1016/S1387-2656\(05\)11007-2](https://doi.org/10.1016/S1387-2656(05)11007-2)
- Montero, V. M. (1986). The interneuronal nature of GABAergic neurons in the lateral geniculate nucleus of the rhesus monkey: a combined HRP and GABA-immunocytochemical study. *Experimental Brain Research*, 64(3), 615–622.  
<https://doi.org/10.1007/BF00340502>
- Moore, G., Jager, P., Delogu, A., Schultz, S., & Brickley, S. (2018). Cell Counting in Targeted Nuclei of Whole Brain Two-Photon Image Data. *Biophotonics Congress: Biomedical Optics Congress 2018 (Microscopy/Translational/Brain/OTS)*, JTU3A.48.  
<https://doi.org/10.1364/TRANSLATIONAL.2018.JTU3A.48>
- Morgan, J. L., Berger, D. R., Wetzel, A. W., & Lichtman, J. W. (2016). The Fuzzy Logic of

- Network Connectivity in Mouse Visual Thalamus. *Cell*, 165(1), 192–206.  
<https://doi.org/10.1016/j.cell.2016.02.033>
- Morgan, J. L., & Lichtman, J. W. (2020). An Individual Interneuron Participates in Many Kinds of Inhibition and Innervates Much of the Mouse Visual Thalamus. *Neuron*, 106(3).  
<https://doi.org/10.1016/j.neuron.2020.02.001>
- Mori, S., Crain, B. J., Chacko, V. P., & Van Zijl, P. C. M. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2), 265–269. [https://doi.org/10.1002/1531-8249\(199902\)45:2<265::AID-ANA21>3.0.CO;2-3](https://doi.org/10.1002/1531-8249(199902)45:2<265::AID-ANA21>3.0.CO;2-3)
- Mulder, I. A., Khmelinskii, A., Dzyubachyk, O., de Jong, S., Rieff, N., Wermer, M. J. H., Hoehn, M., Lelieveldt, B. P. F., & van den Maagdenberg, A. M. J. M. (2017). Automated ischemic lesion segmentation in MRI mouse brain data after transient middle cerebral artery occlusion. *Frontiers in Neuroinformatics*, 11(January), 1–15.  
<https://doi.org/10.3389/fninf.2017.00003>
- Müller-Reichert, T., Mäntler, J., Srayko, M., & O'Toole, E. (2008). Electron microscopy of the early *Caenorhabditis elegans* embryo. *Journal of Microscopy*, 230(2), 297–307.  
<https://doi.org/10.1111/j.1365-2818.2008.01985.x>
- Murman, D. L. (2015). The Impact of Age on Cognition. *Seminars in Hearing*, 36(3), 111–121.  
<https://doi.org/10.1055/s-0035-1555115>
- Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 807–814.
- Nakai, J., Ohkura, M., & Imoto, K. (2001). A high signal-to-noise  $Ca^{2+}$  probe composed of a single green fluorescent protein. *Nature Biotechnology*, 19(2), 137–141.  
<https://doi.org/10.1038/84397>
- Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K., Ito, K., Fukuda, H., Schormann, T., & Zilles, K. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing: A PET study. *Brain*, 123(9), 1903–1912. <https://doi.org/10.1093/brain/123.9.1903>
- Navarro, S. J., Trinh, T., Lucas, C. A., Ross, A. J., Waymire, K. G., & MacGregor, G. R. (2012). The C57BL/6J mouse strain background modifies the effect of a mutation in *Bcl2l2*. *G3: Genes, Genomes, Genetics*, 2(1), 99–102. <https://doi.org/10.1534/g3.111.000778>

- Nazib, A., Galloway, J., Fookes, C., & Perrin, D. (2018). Performance of Registration Tools on High-Resolution 3D Brain Images. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2018-July*, 566–569. <https://doi.org/10.1109/EMBC.2018.8512403>
- Neher, E., & Sakmann, B. (1992). The Patch Clamp Technique. In *Scientific American* (Vol. 266, pp. 44–51). Scientific American, a division of Nature America, Inc. <https://doi.org/10.2307/24938980>
- Nestor, S. M., Gibson, E., Gao, F. Q., Kiss, A., & Black, S. E. (2013). A direct morphometric comparison of five labeling protocols for multi-atlas driven automatic segmentation of the hippocampus in Alzheimer’s disease. *NeuroImage*, 66, 50–70. <https://doi.org/10.1016/j.neuroimage.2012.10.081>
- Niedworok, C. J., Brown, A. P. Y., Jorge Cardoso, M., Osten, P., Ourselin, S., Modat, M., & Margrie, T. W. (2016). *aMAP is a validated pipeline for registration and segmentation of high-resolution mouse brain data*. 7, 11879. <https://doi.org/10.1038/ncomms11879>
- Niimi, K., Kanaseki, T., & Takimoto, T. (1963). The comparative anatomy of the ventral nucleus of the lateral geniculate body in mammals. *The Journal of Comparative Neurology*, 121(3), 313–323. <https://doi.org/10.1002/cne.901210303>
- Nilsson, L.-G. (2003). Memory function in normal aging. *Acta Neurologica Scandinavica. Supplementum*, 179, 7–13. <https://doi.org/10.1034/j.1600-0404.107.s179.5.x>
- Nissl, F. (1894). Ueber eine neue Untersuchungsmethode des Centralorgans zur Feststellung der Localisation der Nervenzellen. *Neurologisches Centralblatt*, 13, 507–508.
- Noakes, C. R., Goto, T., Keller, R., & Periasamy, A. (2001). Does digital deconvolution improve two-photon microscopy in deep tissue imaging? In A. Periasamy & P. T. C. So (Eds.), *Multiphoton Microscopy in the Biomedical Sciences* (Vol. 4262, p. 389). SPIE. <https://doi.org/10.1117/12.424581>
- Ohara, P. T., Sefton, A. J., & Lieberman, A. R. (1980). Mode of termination of afferents from the thalamic reticular nucleus in the dorsal lateral geniculate nucleus of the rat. *Brain Research*, 197(2), 503–506. [https://doi.org/10.1016/0006-8993\(80\)91136-1](https://doi.org/10.1016/0006-8993(80)91136-1)
- Olesen, M. V., Needham, E. K., & Pakkenberg, B. (2017). The optical fractionator technique to estimate cell numbers in a rat model of electroconvulsive therapy. *Journal of Visualized Experiments*, 2017(125). <https://doi.org/10.3791/55737>
- Oron, D., Tal, E., & Silberberg, Y. (2005). Scanningless depth-resolved microscopy. *Optics*



- Express*, 13(5), 1468. <https://doi.org/10.1364/opex.13.001468>
- Oskal, K. R. J., Risdal, M., Janssen, E. A. M., Undersrud, E. S., & Gulsrud, T. O. (2019). A U-net based approach to epidermal tissue segmentation in whole slide histopathological images. *SN Applied Sciences*, 1(7), 1–12. <https://doi.org/10.1007/s42452-019-0694-y>
- Osten, P., & Margrie, T. W. (2013). Mapping brain circuitry with a light microscope. In *Nature Methods* (Vol. 10, Issue 6, pp. 515–523). Nature Publishing Group. <https://doi.org/10.1038/nmeth.2477>
- Otomo, K., Hibi, T., Murata, T., Watanabe, H., Kawakami, R., Nakayama, H., Hasebe, M., & Nemoto, T. (2015). Multi-point scanning two-photon excitation microscopy by utilizing a high-peak-power 1042-nm laser. *Analytical Sciences*, 31(4), 307–313. <https://doi.org/10.2116/analsci.31.307>
- Ouzounov, D. G., Wang, T., Wang, M., Feng, D. D., Horton, N. G., Cruz-Hernández, J. C., Cheng, Y. T., Reimer, J., Tolias, A. S., Nishimura, N., & Xu, C. (2017). In vivo three-photon imaging of activity of Gcamp6-labeled neurons deep in intact mouse brain. *Nature Methods*, 14(4), 388–390. <https://doi.org/10.1038/nmeth.4183>
- Owens, W. A. (1966). Age and mental abilities: A second adult follow-up. In *Journal of Educational Psychology* (Vol. 57, Issue 6, pp. 311–325). American Psychological Association. <https://doi.org/10.1037/h0023962>
- Pan, C., Cai, R., Quacquarelli, F. P., Ghasemigharagoz, A., Lourbopoulos, A., Matryba, P., Plesnila, N., Dichgans, M., Hellal, F., & Ertürk, A. (2016). Shrinkage-mediated imaging of entire organs and organisms using uDISCO. *Nature Methods*, 13(10), 859–867. <https://doi.org/10.1038/nmeth.3964>
- Papagiakoumou, E., Ronzitti, E., & Emiliani, V. (2020). Scanless two-photon excitation with temporal focusing. In *Nature Methods* (Vol. 17, Issue 6, pp. 571–581). Nature Research. <https://doi.org/10.1038/s41592-020-0795-y>
- Pasik, P., Pasik, T., & Hátori, J. (1976). Synapses between interneurons in the lateral geniculate nucleus of monkeys. *Experimental Brain Research*, 25(1), 1–13. <https://doi.org/10.1007/BF00237322>
- Pearce, J. M. S. (2019). The “split brain” and Roger Wolcott Sperry (1913–1994). In *Revue Neurologique* (Vol. 175, Issue 4, pp. 217–220). Elsevier Masson SAS. <https://doi.org/10.1016/j.neurol.2018.07.007>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,

- Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research*, 12, 2825–2830.
- Petráň, M., Hadravský, M., Egger, M. D., & Galambos, R. (1968). Tandem-Scanning Reflected-Light Microscope. *Journal of the Optical Society of America*, 58(5), 661–664. <https://doi.org/10.1364/JOSA.58.000661>
- Petros, T. J., Rebsam, A., & Mason, C. A. (2008). Retinal Axon Growth at the Optic Chiasm: To Cross or Not to Cross. *Annual Review of Neuroscience*, 31(1), 295–315. <https://doi.org/10.1146/annurev.neuro.31.060407.125609>
- Prevedel, R., Verhoef, A. J., Pernía-Andrade, A. J., Weisenburger, S., Huang, B. S., Nöbauer, T., Fernández, A., Delcour, J. E., Golshani, P., Baltuska, A., & Vaziri, A. (2016). Fast volumetric calcium imaging across multiple cortical layers using sculpted light. *Nature Methods*, 13(12), 1021–1028. <https://doi.org/10.1038/nmeth.4040>
- Prince, M., Knapp, M., Guerchet, M., McCrone, P., Prina, M., Comas-Herrera, A., Wittenberg, R., Adelaja, B., Hu, B., King, D., Rehill, A., & Salimkumar, D. (2014). *Dementia UK: Second edition*. <https://doi.org/10.1037/023990>
- Prusti, T., de Bruijne, J. H. J., Brown, A. G. A., Vallenari, A., Babusiaux, C., Bailer-Jones, C. A. L., Bastian, U., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Milligan, D. J., Panem, C., ... Zschocke, S. (2016). The Gaia mission. *A&A*, 595. <https://doi.org/10.1051/0004-6361/201629272>
- Qu, Z., Lin, S.-P., Ju, F.-R., & Liu, L. (2015). The Improved Algorithm of Fast Panorama Stitching for Image Sequence and Reducing the Distortion Errors. *Mathematical Problems in Engineering*, 2015, 428076. <https://doi.org/10.1155/2015/428076>
- 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 for automated ex vivo mouse brain imaging. *Nature Methods*, 9(3), 255–258. <https://doi.org/10.1038/nmeth.1854>
- Ram, N., Rabbitt, P., Stollery, B., & Nesselroade, J. R. (2005). Cognitive performance inconsistency: Intraindividual change and variability. *Psychology and Aging*, 20(4), 623–633. <https://doi.org/10.1037/0882-7974.20.4.623>
- Ramón y Cajal, S. (1909). *Histologie du système nerveux de l'homme & des vertébrés.: Vol. v. 1*. Paris :Maloine,. <https://www.biodiversitylibrary.org/item/103261>

- Ramón Y Cajal, S. (1894). The Croonian Lecture: La Fine Structure des Centres Nerveux. *Proceedings of the Royal Society of London*, 55, 444–468.  
<https://doi.org/10.2307/115494>
- Raz, N., & Rodrigue, K. M. (2006). Differential aging of the brain: Patterns, cognitive correlates and modifiers. *Neuroscience & Biobehavioral Reviews*, 30(6), 730–748.  
<https://doi.org/10.1016/j.neubiorev.2006.07.001>
- Reardon, S. (2017). A giant neuron found wrapped around entire mouse brain. In *Nature* (Vol. 543, Issue 7643, pp. 14–15). <https://doi.org/10.1038/nature.2017.21539>
- 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-N2cδG Strain Enhances Retrograde Synaptic Transfer and Neuronal Viability. *Neuron*, 89(4), 711–724.  
<https://doi.org/10.1016/j.neuron.2016.01.004>
- Reddy, G. D., Kelleher, K., Fink, R., & Saggau, P. (2008). Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. *Nature Neuroscience*, 11(6), 713–720. <https://doi.org/10.1038/nn.2116>
- Renier, N., Adams, E. L., Kirst, C., Wu, Z., Azevedo, R., Kohl, J., Autry, A. E., Kadiri, L., Umadevi Venkataraju, K., Zhou, Y., Wang, V. X., Tang, C. Y., Olsen, O., Dulac, C., Osten, P., & Tessier-Lavigne, M. (2016). Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. *Cell*, 165(7), 1789–1802.  
<https://doi.org/10.1016/j.cell.2016.05.007>
- Renier, N., Wu, Z., Simon, D. J., Yang, J., Ariel, P., & Tessier-Lavigne, M. (2014). iDISCO: A simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell*, 159(4), 896–910. <https://doi.org/10.1016/j.cell.2014.10.010>
- Reynolds, D. (2009). Gaussian Mixture Models. In S. Z. Li & A. Jain (Eds.), *Encyclopedia of Biometrics* (pp. 659–663). Springer US. [https://doi.org/10.1007/978-0-387-73003-5\\_196](https://doi.org/10.1007/978-0-387-73003-5_196)
- Richard, B. C., Kurdakova, A., Baches, S., Bayer, T. A., Weggen, S., & Wirths, O. (2015). Gene dosage dependent aggravation of the neurological phenotype in the 5XFAD mouse model of Alzheimer's disease. *Journal of Alzheimer's Disease*, 45(4), 1223–1236.  
<https://doi.org/10.3233/JAD-143120>
- Richardson, D. S. S., & Lichtman, J. W. W. (2015). Clarifying tissue clearing. *Cell*, 162(2), 246–257. <https://doi.org/10.1016/j.cell.2015.06.067>

- 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>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351, 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychology and Aging*, 20(1), 3–18. <https://doi.org/10.1037/0882-7974.20.1.3>
- Rossi, S., Bartalini, S., Ulivelli, M., Mantovani, A., Di Muro, A., Goracci, A., Castrogiovanni, P., Battistini, N., & Passero, S. (2005). Hypofunctioning of sensory gating mechanisms in patients with obsessive-compulsive disorder. *Biological Psychiatry*, 57(1), 16–20. <https://doi.org/10.1016/j.biopsych.2004.09.023>
- Russell, J. T. (2011). Imaging calcium signals in vivo: A powerful tool in physiology and pharmacology. In *British Journal of Pharmacology* (Vol. 163, Issue 8, pp. 1605–1625). Wiley-Blackwell. <https://doi.org/10.1111/j.1476-5381.2010.00988.x>
- Salat, D. H., Tuch, D. S., Hevelone, N. D., Fischl, B., Corkin, S., Rosas, H. D., & Dale, A. M. (2005). Age-Related Changes in Prefrontal White Matter Measured by Diffusion Tensor Imaging. *Annals of the New York Academy of Sciences*, 1064(1), 37–49. <https://doi.org/10.1196/annals.1340.009>
- Saleeba, C., Dempsey, B., Le, S., Goodchild, A., & McMullan, S. (2019). A student's guide to neural circuit tracing. In *Frontiers in Neuroscience* (Vol. 13, Issue AUG, p. 897). Frontiers Media S.A. <https://doi.org/10.3389/fnins.2019.00897>
- Salthouse, T. A. (2012). Consequences of age-related cognitive declines. *Annual Review of Psychology*, 63, 201–226. <https://doi.org/10.1146/annurev-psych-120710-100328>
- Salthouse, T. A. (2014). Why Are There Different Age Relations in Cross-Sectional and Longitudinal Comparisons of Cognitive Functioning? *Current Directions in Psychological Science*, 23(4), 252–256. <https://doi.org/10.1177/0963721414535212>
- Sauer, B. (1987). Functional expression of the cre-lox site-specific recombination system in

- the yeast *Saccharomyces cerevisiae*. *MOLECULAR AND CELLULAR BIOLOGY*, 7(6), 2087–2096.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–682. <https://doi.org/10.1038/nmeth.2019>
- Schmid, J. S., Bernreuther, C., Nikonenko, A. G., Ling, Z., Mies, G., Hossmann, K. A., Jakovcevski, I., & Schachner, M. (2013). Heterozygosity for the mutated X-chromosome-linked L1 cell adhesion molecule gene leads to increased numbers of neurons and enhanced metabolism in the forebrain of female carrier mice. *Brain Structure and Function*, 218(6), 1375–1390. <https://doi.org/10.1007/s00429-012-0463-9>
- Schnell, S. A., Staines, W. A., & Wessendorf, M. W. (1999). Reduction of lipofuscin-like autofluorescence in fluorescently labeled tissue. *Journal of Histochemistry and Cytochemistry*, 47(6), 719–730. <https://doi.org/10.1177/002215549904700601>
- Schröder, H., Moser, N., Huggenberger, S., Schröder, H., Moser, N., & Huggenberger, S. (2020). The Mouse Thalamus. In *Neuroanatomy of the Mouse* (pp. 171–203). Springer International Publishing. [https://doi.org/10.1007/978-3-030-19898-5\\_8](https://doi.org/10.1007/978-3-030-19898-5_8)
- Schulz, O., Pieper, C., Clever, M., Pfaff, J., Ruhlandt, A., Kehlenbach, R. H., Wouters, F. S., Großhans, J., Bunt, G., & Enderlein, J. (2013). Resolution doubling in fluorescence microscopy with confocal spinning-disk image scanning microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 110(52), 21000–21005. <https://doi.org/10.1073/pnas.1315858110>
- Schwartzman, A. E., Gold, D., Andres, D., Arbuckle, T. Y., & Chaikelson, J. (1987). Stability of intelligence: a 40-year follow-up. *Canadian Journal of Psychology*, 41(2), 244–256. <https://doi.org/10.1037/h0084155>
- Schwarz, G. (1978). Estimating the Dimension of a Model. *Ann. Statist.*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Segebarth, D., Griebel, M., Duerr, A., Collenberg, C. R. von, Martin, C., Fiedler, D., Comeras, L. B., Sah, A., Stein, N., Gupta, R., Sasi, M., Lange, M. D., Tasan, R. O., Singewald, N., Pape, H.-C., Sendtner, M., Flath, C. M., & Blum, R. (2018). DeepFLaSh, a deep learning pipeline for segmentation of fluorescent labels in microscopy images. *BioRxiv*, 473199.

<https://doi.org/10.1101/473199>

- Sherman, S. M. (2004). Interneurons and triadic circuitry of the thalamus. *Trends in Neurosciences*, 27(11), 670–675. <https://doi.org/10.1016/J.TINS.2004.08.003>
- Sherman, S. M. (2005). The thalamic interneuron. In *Dendritic Neurotransmitter Release* (pp. 133–144). Springer US. [https://doi.org/10.1007/0-387-23696-1\\_9](https://doi.org/10.1007/0-387-23696-1_9)
- Sherman, S. M. (2016). Thalamus plays a central role in ongoing cortical functioning. In *Nature Neuroscience* (Vol. 19, Issue 4, pp. 533–541). Nature Publishing Group. <https://doi.org/10.1038/nn.4269>
- Sherman, S. M., & Guillery, R. W. (2006). Exploring the thalamus and its role in cortical function, 2nd ed. In *Exploring the thalamus and its role in cortical function, 2nd ed.* MIT Press.
- Sherman, S. M., & Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 357(1428), 1695–1708. <https://doi.org/10.1098/rstb.2002.1161>
- Shoji, H., Takao, K., Hattori, S., & Miyakawa, T. (2016). Age-related changes in behavior in C57BL/6J mice from young adulthood to middle age. *Molecular Brain*, 9(1). <https://doi.org/10.1186/s13041-016-0191-9>
- Shuvaev, S. A., Lazutkin, A. A., Kedrov, A. V., Anokhin, K. V., Enikolopov, G. N., & Koulakov, A. A. (2017). DALMATIAN: An Algorithm for Automatic Cell Detection and Counting in 3D. *Frontiers in Neuroanatomy*, 11, 117. <https://doi.org/10.3389/fnana.2017.00117>
- Sibarita, J. B. (2005). Deconvolution microscopy. In *Advances in Biochemical Engineering/Biotechnology* (Vol. 95, pp. 201–243). Adv Biochem Eng Biotechnol. <https://doi.org/10.1007/b102215>
- Silvestri, L., Bria, A., Sacconi, L., Iannello, G., & Pavone, F. S. (2012). Confocal light sheet microscopy: micron-scale neuroanatomy of the entire mouse brain. *Optics Express*, 20(18), 20582. <https://doi.org/10.1364/oe.20.020582>
- Silvestri, L., Costantini, I., Sacconi, L., & Pavone, F. S. (2016). Clearing of fixed tissue: a review from a microscopist's perspective. *Journal of Biomedical Optics*, 21(8), 081205. <https://doi.org/10.1117/1.jbo.21.8.081205>
- Simon, M. M., Greenaway, S., White, J. K., Fuchs, H., Gailus-Durner, V., Wells, S., Sorg, T., Wong, K., Bedu, E., Cartwright, E. J., Dacquin, R., Djebali, S., Estabel, J., Graw, J., Ingham, N. J., Jackson, I. J., Lengeling, A., Mandillo, S., Marve, J., ... Brown, S. D. M.

- (2013). A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome Biology*, 14(7), R82. <https://doi.org/10.1186/gb-2013-14-7-r82>
- Singh-Manoux, A., Marmot, M. G., Glymour, M., Sabia, S., Kivimäki, M., & Dugravot, A. (2011). Does cognitive reserve shape cognitive decline? *Annals of Neurology*, 70(2), 296–304. <https://doi.org/10.1002/ana.22391>
- Snowdon, D. A., Greiner, L. H., & Markesbery, W. R. (2000). Linguistic ability in early life and the neuropathology of Alzheimer's disease and cerebrovascular disease: Findings from the Nun Study. *Annals of the New York Academy of Sciences*, 903, 34–38. <https://doi.org/10.1111/j.1749-6632.2000.tb06347.x>
- So, P. T. C., Kim, H., & Kochevar, I. E. (1998). Two-photon deep tissue ex vivo imaging of mouse dermal and subcutaneous structures. *Optics Express*, 3(9), 339. <https://doi.org/10.1364/oe.3.000339>
- Sole-Padulles, C. (2005). Repetitive transcranial magnetic stimulation effects on brain function and cognition among elders with memory dysfunction. A randomized sham-controlled study. *Cerebral Cortex*, 16(10), 1487–1493. <https://doi.org/10.1093/cercor/bhj083>
- Steriade, M. (2005). Sleep, epilepsy and thalamic reticular inhibitory neurons. In *Trends in Neurosciences* (Vol. 28, Issue 6 SPEC. ISS., pp. 317–324). Elsevier Ltd. <https://doi.org/10.1016/j.tins.2005.03.007>
- Stern, Y. (2006). Cognitive reserve and Alzheimer disease. In *Alzheimer Disease and Associated Disorders* (Vol. 20, Issue 2, pp. 112–117). <https://doi.org/10.1097/01.wad.0000213815.20177.19>
- Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of Education and Occupation on the Incidence of Alzheimer's Disease. *JAMA: The Journal of the American Medical Association*, 271(13), 1004–1010. <https://doi.org/10.1001/jama.1994.03510370056032>
- Sun, L., Tang, Y., Yan, K., Yu, J., Zou, Y., Xu, W., Xiao, K., Zhang, Z., Li, W., Wu, B., Hu, Z., Chen, K., Fu, Z. F., Dai, J., & Cao, G. (2019). Differences in neurotropism and neurotoxicity among retrograde viral tracers. *Molecular Neurodegeneration*, 14(1). <https://doi.org/10.1186/s13024-019-0308-6>
- Susaki, E. A., Tainaka, K., Perrin, D., Kishino, F., Tawara, T., Watanabe, T. M., Yokoyama, C., Onoe, H., Eguchi, M., Yamaguchi, S., Abe, T., Kiyonari, H., Shimizu, Y., Miyawaki, A.,

- Yokota, H., & Ueda, H. R. (2014). Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. *Cell*, 157(3), 726–739.  
<https://doi.org/10.1016/j.cell.2014.03.042>
- Susaki, E. A., Tainaka, K., Perrin, D., Yukinaga, H., Kuno, A., & Ueda, H. R. (2015). Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging. *Nature Protocols*, 10(11), 1709–1727. <https://doi.org/10.1038/nprot.2015.085>
- Taghanaki, S. A., Zheng, Y., Zhou, S. K., Georgescu, B., Sharma, P., Xu, D., Comaniciu, D., & Hamarneh, G. (2018). Combo Loss: Handling Input and Output Imbalance in Multi-Organ Segmentation. *Computerized Medical Imaging and Graphics*, 75, 24–33.  
<http://arxiv.org/abs/1805.02798>
- Takasaki, K., Abbasi-Asl, R., & Waters, J. (2020). Superficial bound of the depth limit of two-photon imaging in mouse brain. *ENeuro*, 7(1). <https://doi.org/10.1523/ENEURO.0255-19.2019>
- Tan, C., Guan, Y., Feng, Z., Ni, H., Zhang, Z., Wang, Z., Li, X., Yuan, J., Gong, H., Luo, Q., & Li, A. (2020). DeepBrainSeg: Automated Brain Region Segmentation for Micro-Optical Images With a Convolutional Neural Network. *Frontiers in Neuroscience*, 14, 179.  
<https://doi.org/10.3389/fnins.2020.00179>
- Tarantini, S., Valcarcel-Ares, M. N., Toth, P., Yabluchanskiy, A., Tucsek, Z., Kiss, T., Hertelendy, P., Kinter, M., Ballabh, P., Süle, Z., Farkas, E., Baur, J. A., Sinclair, D. A., Csiszar, A., & Ungvari, Z. (2019). Nicotinamide mononucleotide (NMN) supplementation rescues cerebrovascular endothelial function and neurovascular coupling responses and improves cognitive function in aged mice. *Redox Biology*, 24, 101192. <https://doi.org/https://doi.org/10.1016/j.redox.2019.101192>
- Tarnowski, B. I., Spinale, F. G., & Nicholson, J. H. (1991). DAPI as a useful stain for nuclear quantitation. *Biotechnic and Histochemistry*, 66(6), 296–302.  
<https://doi.org/10.3109/10520299109109990>
- Terribilli, D., Schaufelberger, M. S., Duran, F. L. S., Zanetti, M. V., Curiati, P. K., Menezes, P. R., Scazufca, M., Amaro, E., Leite, C. C., & Busatto, G. F. (2011). Age-related gray matter volume changes in the brain during non-elderly adulthood. *Neurobiology of Aging*, 32(2), 354–368. <https://doi.org/10.1016/j.neurobiolaging.2009.02.008>
- 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. *Neuron*, 92(2), 372–382. <https://doi.org/10.1016/j.neuron.2016.09.021>
- The C. elegans Sequencing Consortium. (1998). Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science*, 282(5396), 2012 LP – 2018. <https://doi.org/10.1126/science.282.5396.2012>
- Thompson, R. H., & Swanson, L. W. (2010). Hypothesis-driven structural connectivity analysis supports network over hierarchical model of brain architecture. *Proceedings of the National Academy of Sciences of the United States of America*, 107(34), 15235–15239. <https://doi.org/10.1073/pnas.1009112107>
- Todorov, M. I., Paetzold, J. C., Schoppe, O., Tetteh, G., Shit, S., Efremov, V., Todorov-Völgyi, K., Düring, M., Dichgans, M., Piraud, M., Menze, B., & Ertürk, A. (2020). Machine learning analysis of whole mouse brain vasculature. *Nature Methods*, 17(4), 442–449. <https://doi.org/10.1038/s41592-020-0792-1>
- Tournier, J.-D., Calamante, F., Gadian, D. G., & Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23(3), 1176–1185. <https://doi.org/10.1016/j.neuroimage.2004.07.037>
- Tsumoto, T., & Suda, K. (1980). Three groups of cortico-geniculate neurons and their distribution in binocular and monocular segments of cat striate cortex. *Journal of Comparative Neurology*, 193(1), 223–236. <https://doi.org/10.1002/cne.901930115>
- Ugolini, G. (2011). Rabies Virus as a Transneuronal Tracer of Neuronal Connections. In *Advances in Virus Research* (Vol. 79). <https://doi.org/10.1016/B978-0-12-387040-7.00010-X>
- United Nations. (2019). *World Population Prospects 2019*.
- Voineskos, A. N., Rajji, T. K., Lobaugh, N. J., Miranda, D., Shenton, M. E., Kennedy, J. L., Pollock, B. G., & Mulsant, B. H. (2012). Age-related decline in white matter tract integrity and cognitive performance: a DTI tractography and structural equation modeling study. *Neurobiology of Aging*, 33(1), 21–34. <https://doi.org/10.1016/j.neurobiolaging.2010.02.009>
- Wahlby, C., Sintorn, I.-M., Erlandsson, F., Borgefors, G., & Bengtsson, E. (2004). Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in

- tissue sections. *Journal of Microscopy*, 215(1), 67–76. <https://doi.org/10.1111/j.0022-2720.2004.01338.x>
- 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, S., Bickford, M. E., van Horn, S. C., Erisir, A., Godwin, D. W., & Sherman, S. M. (2001). Synaptic targets of thalamic reticular nucleus terminals in the visual thalamus of the cat. *The Journal of Comparative Neurology*, 440(4), 321–341. <https://doi.org/10.1002/cne.1389>
- Wang, S., Eisenback, M., Datskovskaia, A., Boyce, M., & Bickford, M. E. (2002). GABAergic pretectal terminals contact GABAergic interneurons in the cat dorsal lateral geniculate nucleus. *Neuroscience Letters*, 323(2), 141–145. [https://doi.org/10.1016/S0304-3940\(01\)02533-2](https://doi.org/10.1016/S0304-3940(01)02533-2)
- Wang, T., Ouzounov, D. G., Wu, C., Horton, N. G., Zhang, B., Wu, C. H., Zhang, Y., Schnitzer, M. J., & Xu, C. (2018). Three-photon imaging of mouse brain structure and function through the intact skull. *Nature Methods*, 15(10), 789–792. <https://doi.org/10.1038/s41592-018-0115-y>
- Wang, X., Sommer, F. T., & Hirsch, J. A. (2011). Inhibitory circuits for visual processing in thalamus. *Current Opinion in Neurobiology*, 21(5), 726–733. <https://doi.org/10.1016/j.conb.2011.06.004>
- Wang, Y., Yau, Y. Y., Perkins-Balding, D., & Thomson, J. G. (2011). Recombinase technology: Applications and possibilities. In *Plant Cell Reports* (Vol. 30, Issue 3, pp. 267–285). Plant Cell Rep. <https://doi.org/10.1007/s00299-010-0938-1>
- Wang, Yun, Ye, M., Kuang, X., Li, Y., & Hu, S. (2018). A simplified morphological classification scheme for pyramidal cells in six layers of primary somatosensory cortex of juvenile rats. *IBRO Reports*, 5, 74–90. <https://doi.org/10.1016/j.ibror.2018.10.001>
- Weber, A. J., Kalil, R. E., & Behan, M. (1989). Synaptic connections between corticogeniculate axons and interneurons in the dorsal lateral geniculate nucleus of the cat. *The Journal of Comparative Neurology*, 289(1), 156–164. <https://doi.org/10.1002/cne.902890113>

- Weber, M., Wu, T., Hanson, J. E., Alam, N. M., Solanoy, H., Ngu, H., Lauffer, B. E., Lin, H. H., Dominguez, S. L., Reeder, J., Tom, J., Steiner, P., Foreman, O., Prusky, G. T., & Levie, K. S. (2015). Cognitive deficits, changes in synaptic function, and brain pathology in a mouse model of normal aging. *ENeuro*, 2(5). <https://doi.org/10.1523/ENEURO.0047-15.2015>
- Whalley, L. J., Starr, J. M., Athawes, R., Hunter, D., Pattie, A., & Deary, I. J. (2000). Childhood mental ability and dementia. *Neurology*, 55(10), 1455–1459. <https://doi.org/10.1212/WNL.55.10.1455>
- 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>
- Woods, R. P., Cherry, S. R., & Mazziotta, J. C. (1992). Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, 16(4), 620–633.
- Xing, J., & Miao, Z. (2007). An improved algorithm on image stitching based on SIFT features. *Second International Conference on Innovative Computing, Information and Control, ICICIC 2007*. <https://doi.org/10.1109/ICICIC.2007.147>
- Xu, C. S., Hayworth, K. J., Lu, Z., Grob, P., Hassan, A. M., García-Cerdán, J. G., Niyogi, K. K., Nogales, E., Weinberg, R. J., & Hess, H. F. (2017). Enhanced FIB-SEM systems for large-volume 3D imaging. *ELife*, 6. <https://doi.org/10.7554/eLife.25916>
- Zeiler, M. D., Krishnan, D., Taylor, G. W., & Fergus, R. (2010). Deconvolutional networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2528–2535. <https://doi.org/10.1109/CVPR.2010.5539957>
- Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45–57. <https://doi.org/10.1109/42.906424>
- Zheng, Z., Lauritzen, S., Perlman, E., Robinson, C., Nichols, M., Milkie, D., Torrens, O., Price, J., Fisher, C., Sharifi, N., Calle-Schuler, S., Kmecova, L., Ali, I., Karsh, B., Trautman, E.,

- Bogovic, J., Hanslovsky, P., Jefferis, G., Kazhdan, M., ... Bock, D. (2017). A Complete Electron Microscopy Volume Of The Brain Of Adult *Drosophila melanogaster*. *BioRxiv*, 140905. <https://doi.org/10.1101/140905>
- Zingg, B., Chou, X. lin, Zhang, Z. gang, Mesik, L., Liang, F., Tao, H. W., & Zhang, L. I. (2017). AAV-Mediated Anterograde Transsynaptic Tagging: Mapping Corticocollicular Input-Defined Neural Pathways for Defense Behaviors. *Neuron*, 93(1), 33–47. <https://doi.org/10.1016/j.neuron.2016.11.045>
- Zingg, B., Hintiryan, H., Gou, L., Song, M. Y., Bay, M., Bienkowski, M. S., Foster, N. N., Yamashita, S., Bowman, I., Toga, A. W., & Dong, H. W. (2014). Neural networks of the mouse neocortex. *Cell*, 156(5), 1096–1111. <https://doi.org/10.1016/j.cell.2014.02.023>
- Zipfel, W. R., Williams, R. M., & Webb, W. W. (2003). Nonlinear magic: Multiphoton microscopy in the biosciences. In *Nature Biotechnology* (Vol. 21, Issue 11, pp. 1369–1377). Nature Publishing Group. <https://doi.org/10.1038/nbt899>