Reverse translational approaches to population neuroimaging: Using rodent models to interpret genetic associations with MRI



Cristiana Tisca

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

Dragilor mei părinţi

To my beloved parents

Acknowledgements

This thesis is the result of an extensive collaborative effort across the Wellcome Centre for Integrative Neuroimaging (WIN).

I am forever grateful to my supervisors, Karla Miller and Aurea Martins-Bach, for their mentorship and continuous support throughout the last 4 years. In 2017, I saw Karla give a talk during a workshop while I was doing a summer research project at UCL. Her research work and presentation skills stood out and, as a budding female researcher, I saw a role model in her. When I started applying for PhDs, I decided to reach out to her. Not only did she get back to me, but she went above and beyond in helping me prepare my application. I am so grateful that she believed in me – this changed the course of my life.

I am thankful to Aurea for her continuous help and support. Before joining Oxford, I did not have extensive practical laboratory experience. From Aurea, I learnt to feel comfortable in all laboratory environments: from operating and troubleshooting an MRI scanner, to planning and conducting very finicky histology experiments. Aurea showed me how to be a multi-faced scientist, thriving at the intersection between physics and biology.

Thank you to WIN's Physics, Analysis and Preclinical Imaging groups. Mohamed Tachrount helped extensively with setting up and troubleshooting our MRI acquisition protocols. Adele Smart made invaluable contributions to the histology component of my project. Amy Howard, Rick Lange, Chaoyue Wang and Benjamin Tendler provided help and advice with the development of the MRI post-processing pipelines. Istvan Huszar and Daniel Kor helped combine the MRI and histology datasets. Steve Smith and Michiel Cottaar advised on the composite modelling of the MRI-histology datasets. Jason Lerch and his team – in particular, Claire Bratley, Lily Qiu and Urs Schuffelgen – provided extremely useful advice and help with conducting experiments during the lockdown period.

Throughout my DPhil, I loved interacting with the wonderful people at WIN, particularly during GLOW meetings, public engagement activities and the regular pub trips. I am very lucky to have such wonderful friends at the centre. Thank you to Daniel, Xi, Thijs, James, Hossein and Amy. I also loved being an active member of Linacre College, meeting people outside my field and making lifelong friendships. Thank you to Jane Hoverd, Linacre's Senior Tutor, for her invaluable support and advice during my time at Linacre.

Early in my DPhil I started suffering from chronic pain. I am beyond grateful for all the medical help I received and continue to receive as I am fighting my way towards a pain-free life.

This thesis is dedicated to my parents. They never got the chance to attend university, but made great sacrifices to make sure that my sister and I will get the best possible education.

And lastly, to the person who was right by my side throughout my DPhil: my partner, Zeev. You are the most precious gift I got from my time in Oxford. Thank you for your love and boundless patience and for always taking care of me. Nothing makes me happier than knowing I'll always have you by my side.

Abstract

Mouse models play an essential role in advancing research in humans. The genetic similarity between humans and mice means we can measure and compare equivalent phenotypes across species. More importantly, through invasive studies which we cannot conduct in humans, mouse models can provide a mechanistic insight into the molecular and cellular drivers of the phenomena of interest.

This thesis proposes to investigate key findings from the UK Biobank population neuroimaging studies using mouse models. Researchers have identified novel links between imaging-derived phenotypes and genetics. Key results of these studies showed associations between genetic variants in the *VCAN* and *BCAN* genes and multi-modal MRI metrics in white matter. *VCAN* and *BCAN* encode extracellular matrix proteins. The biological drivers of these associations are not obvious, in that we do not expect the extracellular matrix to have a strong influence on the identified MRI signals.

To investigate the biological mechanisms driving the observed MRI signal changes, we conducted multi-modal *ex-vivo* MRI and histology studies in *Vcan* and *Bcan* mouse models.

We first developed new MRI and histology protocols, neither of which were established at the start of our studies. We implemented new multi-modal *ex-vivo* MRI protocols and post-processing pipelines, working in a new Bruker 7T scanning facility. We subsequently optimised immunohistochemistry processing and staining protocols to yield high-quality, multi-stain histology data. This data is registered to MRI using bespoke tools and pipelines developed in-house and processed to extract quantitative metrics which can be directly linked to MRI metrics at the MRI voxel level.

Our *ex-vivo* MRI results showed that some of the *Vcan* and *Bcan* homozygous mice exhibited differences in MRI phenotypes compared to wild-type controls. In particular, male homozygous *Bcan* mice showed consistent MRI phenotypes across multiple

metrics when compared to male wild-types. These findings confirmed that changes in the *Vcan* and *Bcan* genes lead to changes in brain imaging phenotypes. Our preliminary histology analyses reveal some emerging trends in the histology data. In male *Bcan* mice, we observed differences in neurofilament content which would be consistent with the observed changes in MRI metrics. These trends may prove informative once we have completed the histological analysis on the full cohorts.

Institutional permissions

Mouse brain samples used in this thesis were collected by collaborators in the United Kingdom and Canada. Perfusions were performed following the guidelines of the UK Home Office legislation (Animals Scientific Procedures Act 1986), with the approval of the relevant ethics committees and/or the United Kingdom Home Office and the Hospital for Sick Children (Toronto, Ontario). Institutional approval was not required for experiments involving *ex-vivo* mouse tissue.

List of abbreviations

ADC	Apparent diffusion coefficient
AK	Apparent kurtosis
AP	Anterior-posterior
ATS	Animal transport system
BW	Bandwidth
ChABC	Chondroitinase ABC
CSF	Cerebrospinal fluid
CSPG	Chondroitin sulphate proteoglycans
DAB	3,3α-diaminobenzidine
DEC	Directionally-encoded colour
DNA	Deoxyribonucleic acid
DKI	Diffusion kurtosis imaging
DOF	Degrees of freedom
DTI	Diffusion tensor imaging
dMRI	Diffusion-weighted MRI
ECM	Extracellular matrix
EPI	Echo-planar imaging
eQTL	Expression quantitative trait loci
FA	Fractional anisotropy
FDT	FMRIB's Diffusion Toolbox
FFPE	Formalin-fixed, paraffin-embedded
FOD	Fibre orientation dispersion
FoV	Field of view
FLAIR	Fluid-attenuated inversion recovery
FSL	FMRIB Software Library
FT	Fourier transform
FWER	Family-wise error rate
Gd	Gadolinium
GE	Gradient-echo
GFAP	Glial fibrillary acidic protein
GRE	Gradient echo
GWAS	Genome-wide association studies
HIER	Heat-induced epitope retrieval
HOM	Homozygous
ICVF	Intra-cellular volume fraction
IDP	Image-derived phenotype
IHC	Immunohistochemistry
iLSQR	Iterative least-squares
ISOVF	Isotropic volume fraction
LD	Linkage disequilibrium
LR	Left-right
MAF	Minor allele frequency
MB	Multiband

MBP	Myelin basic protein
MPRAGE	Magnetisation-prepared rapid acquisition gradient-echo
MRI	Magnetic resonance imaging
MRS	Magnetic resonance spectroscopy
MD	Mean diffusivity
NA	Number of averages
NF	Neurofilaments
NODDI	Neurite orientation dispersion and density imaging
N-sg	Non-staggered
OD	Orientation dispersion
ODI	Orientation dispersion index
PBS	Phosphate-buffered saline
PF	Partial Fourier
PFA	Paraformaldehyde
PLP	Proteolipid protein
PNN	Perineuronal nets
PV	ParaVision
QC	Quality control
QSM	Quantitative susceptibility mapping
RARE	Rapid acquisition with relaxation enhancement
RF	Radiofrequency
RNA	Ribonucleic acid
ROI	Region of interest
SAF	Stain area fraction
SE	Spin-echo
Sg	Staggered
SI	Superior-inferior
SMT	Spherical mean technique
SNP	Single nucleotide polymorphism
SNR	Signal-to-noise ratio
SNV	Single nucleotide variant
STI	Structure tensor imaging
TE	Echo time
TE1	First echo time
TI	Inversion time
TIRL	Tensor image registration library
TR	Repetition time
TSE	Turbo spin-echo
vSHARP	Variable-kernel sophisticated harmonic artifact reduction for phase data
WT	Wild-type

Contents

1.Introduction – page 14

2. Key biological concepts and theories – page 16
2.1. Basic genetics terminology – page 17
2.2. Genetic epidemiology and genome-wide association studies – page 20
2.3. Post-GWAS *reverse* translational approaches using mouse models – page 23
2.4. Brain micro- and macrostructure across species – page 25
2.5. The brain extracellular matrix – page 27
2.5.1. Versican – page 28
2.5.2. Brevican – page 29
2.5.3. Imaging the extracellular matrix – page 30

3. Imaging principles: Magnetic resonance imaging and histology – page 31
3.1. Fundamentals of magnetic resonance imaging – page 33
3.1.1. Magnetic resonance – page 33
3.1.2. MR relaxation times – page 34
3.1.3. MR image formation – page 37
3.1.4. Pulse sequences – page 41
3.1.5. Image contrast – page 43
3.1.5.1. Diffusion contrast – page 45
3.1.5.2. Contrast enhancement – page 48
3.1.6. Signal-to-noise ratio – page 48
3.1.7. Structural MRI processing and analysis – page 49
3.1.8. *Ex-vivo* MRI: Imaging considerations – page 49
3.2. Imaging microstructure using quantitative MRI – page 50
3.2.1. Diffusion-weighted MRI – page 51
3.2.1.1. Data processing – page 51
3.2.1.2. Data modelling – page 51
3.2.1.2.1. Diffusion tensor imaging – page 52
3.2.1.2.2. The NODDI model – page 53
3.2.2. T_2^* relaxometry: Data processing and fitting – page 55
3.2.3. Quantitative susceptibility mapping: Data processing and fitting – page 56
3.3. Imaging genetics – page 58
3.4. Basics of histology – page 61
3.5. MRI-histology data fusion – page 64
3.5.1. MRI-histology registration – page 64
3.5.2. Extraction of quantitative metrics – page 65

4. Multi-modal MRI in *ex vivo* mouse brains: Development of acquisition and post-processing protocols comparable to UK Biobank – page 66
4.1. Introduction – page 70
4.2. Data acquisition – page 71
4.2.1. Hardware considerations – page 71
4.2.2. Protocol development – page 72
4.2.2.1. Structural T_2-weighted MRI – page 74
4.2.2.2. Diffusion-weighted MRI – page 76
4.2.2.3. Multi-echo gradient-echo – page 78
4.3. Sample preparation and scanning – page 80
4.3.1. Perfusion fixation – page 80
4.3.2. Sample loading – page 80
4.3.3. Sample tests – page 81
4.3.4. Data acquisition – page 83
4.4. Data processing – page 83
4.4.1. T_2-weighted MRI – page 83
4.4.2. Diffusion-weighted MRI – page 84
4.4.3. R_2^*-mapping and quantitative susceptibility mapping – page 88
4.4.3.1. Artifact correction – page 88
4.4.3.2. R_2^* processing pipeline – page 89
4.4.3.3. Quantitative susceptibility mapping processing pipeline – page 89
4.5. Co-registration – page 93
4.6. Quality control – page 93
4.7. Data open-access – page 94
4.8. Discussion – page 94
4.9. Conclusion – page 95

5. Histology in mouse brains prepared for *ex vivo* MRI: An optimised framework for data acquisition and processing – page 96
5.1. Introduction – page 99
5.2. Experimental design considerations – page 100
5.2.1. Study hypotheses – page 100
5.2.2. Selection of immunostains – page 101
5.2.3. Data acquisition – page 102
5.3. Histological data acquisition: methodology and results – page 104
5.3.1. Brain dissection – page 104
5.3.2. Brain cutting – page 105
5.3.3. Tissue processing and embedding – page 105
5.3.4. Tissue slicing – page 108
5.3.5. Immunohistochemistry – page 110
5.3.5.1. Myelin PLP and neurofilaments immunostaining – page 113
5.3.5.2. Optimisation of the versican and brevican immunostains – page 116

5.3.5.2.1. Versican – page 116
5.3.5.2.2. Brevican – page 123
5.4. Quality control – page 127
5.5. Histological data processing: methodology and results – page 128
5.5.1. Image registration – page 128
5.5.1.1. MRI-histology registration – page 128
5.5.1.2. Histology-histology registration – page 131
5.5.1.3. MRI-histology framework: registration chain – page 131
5.5.2. Quantitative staining estimates – page 133
5.6. Data open access – page 137
5.7. Discussion – page 137
5.8. Conclusions – page 138

6. Phenotyping of the *Vcan* and *Bcan* mouse models using multi-modal *ex-vivo* MRI and histology – page 139
6.1. Introduction – page 141
6.2. Materials and methods – page 144
6.2.1. The *Vcan* mouse model – page 144
6.2.2. The *Bcan* mouse model – page 145
6.2.3. Sample preparation, data acquisition and processing – page 146
6.2.4. Statistical analysis – page 149
6.2.4.1. MRI – page 149
6.2.4.2. MRI-histology analysis – page 150
6.2.4.2.1. Simple correlation – page 150
6.2.4.2.2. Composite modelling – page 150
6.3. Results – page 153
6.3.1. MRI – page 153
6.3.1.1. Volumetric analysis – page 153
6.3.1.2. Grey-white matter ratio – page 153
6.3.1.3. Diffusion-weighted MRI – page 154
6.3.1.4. R_2^* and QSM – page 156
6.3.2. MRI-histology analysis – page 158
6.3.2.1. Qualitative assessment of the histology data – page 158
6.3.2.2. MRI-histology voxel-wise correlations – page 163
6.3.2.3. Composite modelling – page 170
6.4. Discussion – page 171
6.5. Conclusions – page 175

7. Final remarks and future directions – page 176
7.1. Future work – page 177
7.1.1. Structure tensor imaging – page 177
7.1.2. Subsequent *ex-vivo* MRI data acquisition – page 179

7.1.3. Subsequent histology data acquisition – page 179
7.1.4. Composite modelling – page 179
7.1.5. Long-term research plans – page 180
7.2. Final remarks – page 181

Appendix A: A protocol for high-resolution multi-modal MRI phenotyping of *ex-vivo* mouse brains – Data acquisition and processing – page 182

Appendix B: Optimisation of the anti-versican and anti-brevican immunohistochemistry staining in mouse brains prepared for *ex-vivo* MRI – page 195

Bibliography – page 198

'Since when', he asked,
'Are the first and last line of any poem
Where the poem begins and ends?'
from 'The Fragment'
by Seamus Heaney

Chapter 1: Introduction

Traditionally, biological research is conducted in a "benchtop-to-bedside" direction. This research paradigm faces limitation at the "bedside": due to regulations and ethical constraints, only a limited number of research questions can be answered in humans. Reverse translational approaches – "bedside-to-benchtop" – begin with findings in humans and seek to unravel the drivers of the observed changes in the laboratory, often using animal models (Shakhnovich, 2018). These animal models, in particular mouse models, allow us to replicate experimental discoveries in humans and build on these findings using more invasive approaches.

This thesis employs this "reverse translation" research paradigm to explore the biological significance of key findings from the UK Biobank neuroimaging study, by studying mouse models of relevant genes using both MRI and histology.

The thesis begins with two introductory chapters. **Chapter 2** defines the essential biological concepts relevant to the understanding of the key results from UK Biobank genome-wide association studies. It also introduces the usefulness of reverse translational approaches to explore findings from population studies. It concludes with an overview of brain structure, including the extracellular matrix.

Chapter 3 introduces fundamental MRI concepts, focusing on diffusion-weighted MRI (dMRI), R_2^* and quantitative susceptibility mapping (QSM). It delves into the UK Biobank genome-wide association studies' findings that motivated this research work: notable associations between various MRI metrics and genetic variants in the *VCAN* and *BCAN* genes. Subsequently, it introduces basic histology concepts as well as

important considerations when conducting MRI-histology data fusion: registering the two modalities and extracting reliable quantitative histology metrics.

Chapter 4 presents the development of multi-modal MRI acquisition protocols and post-processing pipelines for *ex-vivo* mouse brain samples. Firstly, we describe the optimisation of the protocols for T_2-weighted structural MRI, dMRI and multi-echo gradient echo sequences. Consequently, we detail the development of dMRI, R_2^*-mapping and QSM post-processing pipelines.

Chapter 5 presents our optimised protocol for formalin-fixed, paraffin-embedded immunohistochemistry in mouse brains prepared for *ex-vivo* MRI. We highlight the challenges we addressed to obtain high-quality brain histological sections. Once stained, these slides can be easily registered to MRI and processed to extract a quantitative histological metric (stain area fraction). **Chapter 5** also highlights our experimental considerations when acquiring the histology datasets for our *Vcan* and *Bcan* mouse cohorts.

Finally, **Chapter 6** presents our on-going work in characterising the *Vcan* and *Bcan* mouse models using the multi-modal MRI and multi-stain histology protocols described in the previous chapters. Mice from both the *Vcan* and *Bcan* cohorts demonstrated phenotypes similar to those present in the UK Biobank study, confirming our hypothesis that alterations in the *Vcan* and *Bcan* genes will drive changes in neuroimaging phenotypes. When analysing this preliminary multi-stain histology data, we observed some emerging trends which would explain the MRI signal changes.

In **Chapter 7**, we highlight our research plans for on-going post-doctoral work, including subsequent analyses of the existing data, as well as plans for acquiring additional MRI and histology data.

Chapter 2: Key biological concepts and theories

Context

This chapter is structured in two parts. The first half will introduce basic genetic terminology and genome-wide association studies, discussing their relevance to population imaging and limitations that motivate the work presented in this thesis. The UK Biobank population imaging study will also be described. The second half of this chapter will discuss neuroanatomy, from the cellular level to macrostructure, with a focus on the brain's extracellular matrix. Where necessary, parallels will be drawn between humans and mouse models, laying the groundwork for the experimental considerations and discussion sections outlined in subsequent thesis chapters.

The genetics information in this introductory chapter is based on the textbook *Molecular Biology of the Cell*, 6th edition, by Alberts and colleagues (Alberts *et al.*, 2017).

2.1. Basic genetics terminology

The genome defines the set of genetic instructions found inside every single cell of our body. All living organisms, including unicellular organisms, rodents and primates, operate under the guidelines of this genetic makeup, which regulate their cellular components and mechanisms and dictate traits ranging from eye colour to disease susceptibility.

The genome contains nucleotide sequences of deoxyribonucleic acid (DNA), *"the language of life"* (Doudna and Sternberg, 2018). The DNA is a macromolecule that contains the genetic code. This code is specified by sequences of four different nucleotides. The DNA is organised in two parallel strands that form a sequence of pairs: adenine (A) pairs exclusively with thymine (T), guanine (G) pairs exclusively with cytosine (C). DNA is organised into structures called chromosomes, found in the nucleus of each cell.

DNA information is encoded into ribonucleic acid (RNA) through DNA transcription. This is followed by translation, which results in the formation of proteins using the information from the RNA molecules. Proteins are large molecules playing essential roles in an organism. These functions are so critical that the lack of certain proteins is a guarantee of the organism's unviability. 22 different types of genetically-encoded amino acids constitute the building blocks of proteins: 20 amino-acids are from the standard genetic code and 2 additional amino-acids can be incorporated by special translation mechanisms.

This two-step process – transcription followed by translation – is known as the central dogma of molecular biology and is conserved across species. Its discovery, with essential contributions from Francis Crick, James Watson and Rosalind Franklin, represents one of the most important scientific achievements in history.

A sequence of DNA base-pairs that encodes information for functional products is defined as a gene (Kampourakis, 2017). The RNA molecule which gets translated into a protein is one of these functional products. Not all parts of a gene get translated into proteins: an exon is a nucleotide sequence which will constitute the final functional RNA produced during transcription, whereas an intron represents a region that will be removed from the RNA molecule (**Figure 2.1**) before translation.

However, that does not mean that introns do not serve a role. Although still intensely researched, introns are known to be vital in regulating the rate at which the information from genes is converted into proteins, known as genetic expression. There are various other important components of genetic expression, such as expression quantitative trait loci (eQTLs), DNA sequences which affect the variation in genetic expression of certain genes. eQTLs can be located in the proximity of the influenced gene or even on a distinct chromosome (Nica and Dermitzakis, 2013). Within chromosomes, the locations of genes and genetic markers are known as “genetic loci”.

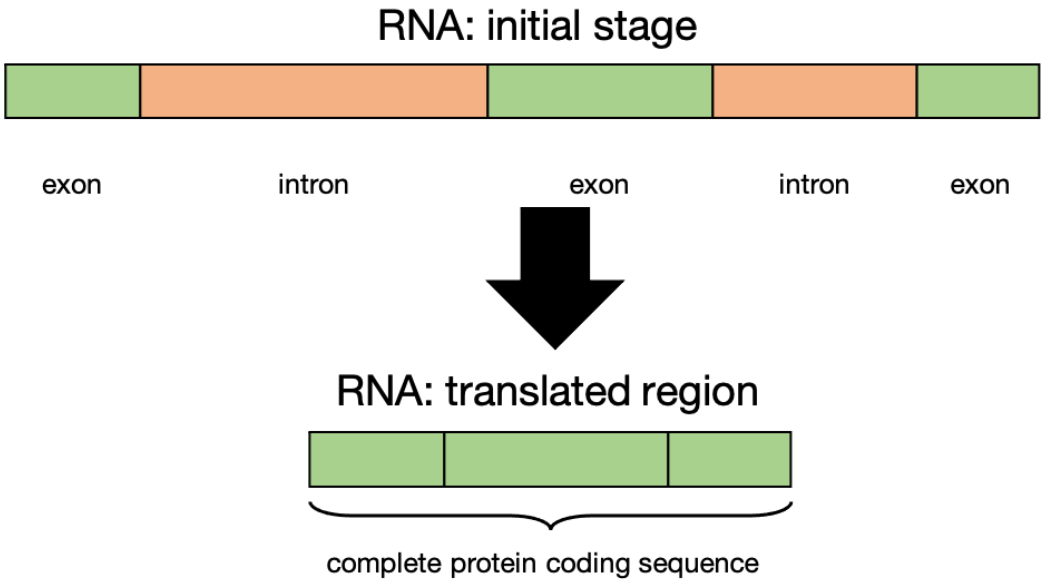

Figure 2.1. Graphical representation of exons and introns within a gene.

The human DNA has about 3.1 billion base-pairs. It is estimated that there are about 20,000 protein-encoding genes in the human genome. Incredibly, 98% of our genome is composed of non-coding DNA, which does not contain information to make proteins. Comparatively, a mouse genome has about 2.7 million base-pairs and 22,000 protein-coding genes (Breschi, Gingeras and Guigo, 2019).

A single gene can provide information for making more than one protein. These different protein "isoforms" are obtained through a process called alternative splicing. The isoforms can have various roles and functions, ultimately influencing multiple characteristics of an organism.

Genetic variation represents differences in DNA sequences across individuals. When this variation occurs at the level of a single nucleotide, it is known as a single nucleotide variant (SNV). Here, a single base-pair is replaced with another, for instance a C-G pair is replaced with a A-T pair (**Figure 2.2**). A single nucleotide polymorphism (SNP) is an SNV that occurs in a substantial fraction of a population, larger than 1% (Manolio *et al.*, 2009). SNPs are also known as common genetic variants. Current DNA sequencing technology allows for measuring of genetic variation at the nucleotide level in individual organisms, and therefore the identification of SNPs in individuals.

The different versions of a genetic sequence at a particular location in the genome are called "alleles". This genetic sequence can be a gene or a smaller sequence of nucleotides which might include a single SNP. At a given location where a SNP is observed, there may be two or more alleles represented by alternative bases. For instance, a SNP could have two versions, represented by the bases C and T, present in 95% and 5% of the population, respectively. It is useful to introduce the concept of major and minor alleles here: the "major allele" is the more common version of the variant (in the example above it would be C), the minor allele is the less common one (T). The "minor allele frequency" (MAF; 5% in the example above) is the frequency at which the minor allele occurs in a population.

A “reference allele” is the version of a genetic sequence found in the reference genome, a digital database assembled by scientists as an example genome. An “alternative allele” refers to any variant other than the reference. The reference allele can be either the major or minor allele. When alleles from different loci associate non-randomly, it is said that they are in “linkage disequilibrium” (LD) (Slatkin, 2008).

Finally, a phenotype represents a trait of an individual. Some phenotypes can be determined by the genotype, the genetic information encoding that particular trait. Phenotypes can be categorical (for example, handedness) or continuous (for example, brain volume). SNPs are key contributors to phenotypic variance; however, their influence over phenotypes is typically subtle.

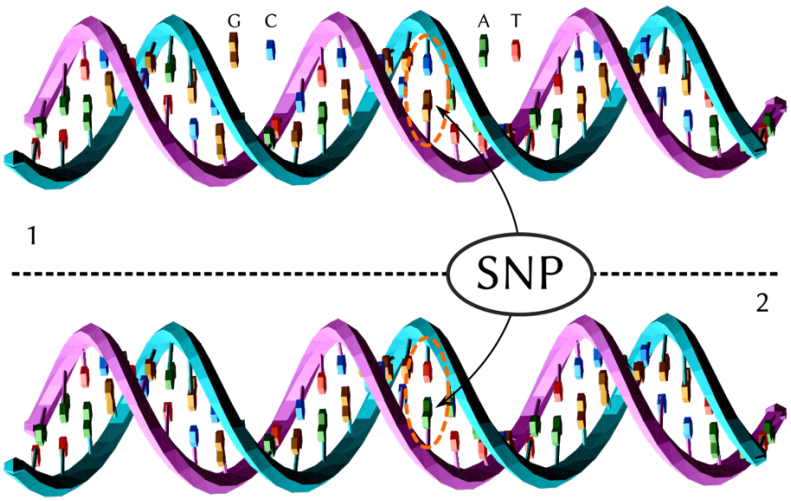

Figure 2.2. Representation of a single nucleotide polymorphism (SNP) between two genomes (1 and 2). From: https://commons.wikimedia.org/wiki/File:Dna-SNP.svg, distributed under GNU Free Documentation License.

2.2. Genetic epidemiology and genome-wide association studies

Genetic epidemiology investigates the genetic cause of health and disease in populations (Burton, Tobin and Hopper, 2005). Genome-wide association studies (GWAS) look for associations between genetic variations, such as SNPs, and phenotypes. GWAS are hypothesis-free, searching large fractions of the genome to

discover new associations without *a priori* assumptions. The simplest GWAS includes a logistic regression for every SNP in the genome (Scelsi, 2020):

$$y = \beta_0 + \beta_1 \times SNP + \sum_{i=2}^{m} \beta_i \times COV_i \quad (2.1)$$

Here, y is a binary vector signifying phenotype status, e.g., disease status (present or absent) for N individuals. The regressor, SNP, can have values of 0, 1 or 2, representing the number of minor allele copies carried by the individual. β_0 is the estimated intercept, β_1 is the estimated regression coefficient relating the SNP value to y, $\beta_{i=2,\dots,m}$ is the vector of the estimated regression coefficients associated to COV_i, an $N \times (m-2)$ matrix representing confounding variables. Logistic regression allows for the calculation of the regression coefficients, by estimating the sigmoid curve that is the best fit to the data. A statistically significant association is one where the β_1 coefficient differs from 0 at the significance level of interest.

Due to the incredibly high number of association tests that are involved in a GWAS, a very large cohort is usually needed to reveal statistically significant results. To avoid false discovery, one needs to perform "multiple comparison correction", such as the Bonferroni method, which divides the desired statistical significance threshold by the total number of tests. There are approximately 1 million independent common genetic variants, which results in a Bonferroni multiple comparison threshold of $p < 5 \times 10^{-8}$ (Uffelmann *et al.*, 2021). Alternatively, the analysis can be restricted to candidate genomic regions. If interested in continuous traits, the GWAS model can be easily modified by using linear regression instead. A statistically significant association between a SNP and a phenotype is often referred to as a "hit".

UK Biobank (Sudlow *et al.*, 2015) is one of the largest epidemiological studies in history, which was specifically designed to enable GWAS to investigate a broad range of health conditions. UK Biobank collects genetic and health data from 500,000 adult participants from the United Kingdom. The health data includes blood samples,

physiological measurements, and life factor questionnaires. Imaging data from the brain, abdomen and the heart are being acquired in a subset of 100,000 participants, making UK Biobank the largest imaging study ever conducted. The first imaging GWAS analysed UK Biobank data from around 11,000 subjects, revealing numerous hits that have never been described before (Elliott *et al.*, 2018). More details about this study are presented in **Chapter 3**.

2.3. Post-GWAS *reverse* translational approaches using mouse models

A central drawback of GWAS is that they do not allow the identification of biological mechanisms responsible for the revealed associations (Tam *et al.*, 2019). Most association signals are in non-coding regions of the genome, which may be a region enhancing gene expression (e.g., eQTL) or in linkage disequilibrium with SNPs in coding regions of the gene. Therefore, to identify and functionally characterise these causal mechanisms, follow-up experiments with well-defined hypotheses are needed.

In many cases, we can better understand the underlying biological mechanisms driving GWAS hits using genetically modified animal models. Animal models can have a particular gene removed (known as a *knock-out* model) or inserted from another organism (*knock-in*) or can have genetic modifications that impair or enhance the function of the resulting protein. Knock-out animals where both copies of the gene from both chromosomes are removed in all tissues are known as “homozygous”. The control animals that they are usually compared to, their “wild-type” littermates, have no genetic modification and are bred and raised under identical conditions. Animal models enable us to gain mechanistic insight into the genetic drivers of phenotypic change. They simultaneously allow for studies investigating levels of molecular compounds, microscopy investigations looking at the distribution of proteins in tissue and at cellular morphology, as well as macroscopic evaluations of phenotypical changes, such as anatomical and behavioural measurements.

The most popular choice of animal model in post-GWAS studies is by far the mouse. This choice is motivated by its genetic similarity to humans, with 75% of genes having equivalents in the human genome (Ermann and Glimcher, 2012), followed by the availability of its complete genomic sequence (Waterston *et al.*, 2002), genetically defined strains (Svenson *et al.*, 2007), extensive genetic manipulation techniques (allowing the removal or insertion of a gene, the reduction of gene expression, etc.) (Gurumurthy and Kent Lloyd, 2019), ease of breeding, ability to control breeding and phenotyping and the availability of a wide array of phenotypic tests (Crawley, 2008).

Inbred mice of the same genetic background (e.g., C57BL/6 or BALB/c) are usually used in these studies and are obtained by breeding littermates for at least 20 generations, ensuring genetic homogeneity. This means that members of the strain are approximately genetically identical. Furthermore, mouse models are the most common option used in neuroscience to replicate findings in humans, principally due to the equivalent micro- and macrostructure of its nervous system, and comparable physiology.

There are numerous examples in literature successfully using mouse models to explore the causality of GWAS hits. For example, Sur et al (Sur *et al.*, 2012) used a knock-out mouse model to confirm that a region around a SNP in a gene associated with increased cancer risk has direct effects on tumorigenesis. Mague and colleagues (Mague *et al.*, 2009) describe a mouse model of a human-equivalent SNP associated with drug addiction phenotypes which displays several similarities in behavioural phenotypes to humans carrying the SNP. Their studies revealed sex-specific differences in the studied mice, which were previously unreported in humans, warranting follow-up analyses. Hernandez Cordero and others (Hernandez Cordero *et al.*, 2019) also found equivalent phenotypes when investigating myogenesis using GWAS in humans and mice.

Perhaps some of the most exciting new GWAS findings are in neuroscience, particularly due to advances in imaging methods, such as magnetic resonance imaging (MRI) (Elliott et al., 2018a; Smith et al., 2021; Wang et al., 2021). These studies revealed novel associations between SNPs and imaging-derived phenotypes. These results will be described in more depth in the next chapter, after introducing the fundamentals of the MRI methods they use. Follow-up studies using mouse models present great potential to elucidate the biological drivers of the findings. These studies are missing from the literature. The work presented in this thesis aims to fill this research gap.

2.4. Brain micro- and macrostructure across species

2,400 years ago, Hippocrates designated the brain as the organ responsible with "operations of the understanding". It took millennia to start examining how the structure of the brain relates to its function. About 120 years ago, neuroscientists came up with the neuron doctrine, chiefly because of pioneering microscopy work by Santiago Ramón y Cajal. This states that the nervous system, like the rest of the organism, is made up of cells. We now know significantly more about the information-conducting and processing cells of the nervous system, the neurons, and the other cells, glia, which have extensive functional and neuroprotective roles.

Neurons differ from other cells in the body in both shape and function. They conduct information in the brain via electrical impulses (action potentials), as well as chemical signals. This information gets passed from neuron to neuron at the synapse. Neurons typically have a large projection emerging from the cell body which acts as an information output channel – the axon – and dendrites, small projections surrounding the cell body which collect and process information from other neurons synapsing onto them. Both axons and dendrites are known as neurites. The other major cell type that is unique to the nervous system are glia. Glial subtypes include oligodendrocytes, which make up myelin sheets wrapping around neuronal axons, improving the conduction of information via electrical impulses; astrocytes, responsible, among many other functions, for providing nutrients to neurons by linking to blood vessels; and microglia, triggered during the immune response to injury.

Brain tissue can be classified into white and grey matter when viewed macroscopically. White matter includes brain-wide tracts, bundles of myelinated axons tightly grouped together. Other cell bodies, especially oligodendrocytes, can be found in white matter. Grey matter includes neuronal cell bodies, their dendrites, other glial cells, and axons projecting to nearby brain areas.

Brain structure is similar across species despite dramatic changes in volume: the human brain volume is about 1200cm^3, while the mouse brain volume is around

0.5cm^3. Brain micro- and macrostructure are equivalent across the two species, with homologous brain areas and cellular architecture present. A major difference between human and rodent brains is in morphology: the human brain presents a gyrification pattern across the cortex which maximises surface area, whereas mouse brains are essentially smooth. While equivalent areas exist across brains of different species for many basic functions, such as the visual or motor cortex, these regions have proportionally different volumes compared to the rest of the brain across species. For instance, the olfactory bulb is proportionally larger in mice than humans, as mice rely heavily on their sense of smell. The relative proportion of white matter in the human and mouse brains is also strikingly dissimilar: a relatively small number of thin white matter tracts are present in the mouse brain, most of which connect between hemispheres. Finally, not all brain regions found in humans have homologous regions in other species, with the general pattern that more complex behaviours are subserved by emergent specialist brain regions and associated white matter pathways.

2.5. The brain extracellular matrix

Two of the strongest hits from the first UK Biobank imaging GWAS were in genes encoding extracellular matrix (ECM) proteins, *VCAN* and *BCAN* (Elliott *et al.*, 2018). In this thesis, we characterise mouse models with modifications in the *Vcan* and *Bcan* genes using MRI and histology. This section introduces the brain ECM and discusses the function of the versican and brevican proteins, encoded by the *VCAN* and *BCAN* genes, respectively.

All mammalian cells are surrounded by the ECM, a material that provides structural stability, influences cellular development and interaction, and minimises the diffusion of soluble and membrane-associated molecules. The brain ECM occupies approximately 20% of total brain volume in the adult brain (Nicholson and Syková, 1998). Its highly interacting molecular components, produced by neurons and glia, include chondroitin sulphate proteoglycans (CSPGs) (such as versican and brevican), hyaluronan (also known as hyaluronic acid) and other link proteins. The brain ECM changes during development, disease progression and experience.

CSPGs are some of the largest and most complex molecular structures surrounding mammalian cells (Iozzo, 1998) and share common structural and functional similarities. All CSPGs are negatively charged, attracting water molecules (Wight *et al.*, 2014), which affects diffusion in the ECM environment. While they can be found throughout the diffuse brain ECM, they are also concentrated in complex ECM structures. Around some neuronal cell bodies (such as inhibitory neurons), they form essential components of perineuronal nets (PNNs), a honeycomb-shaped mesh structure. PNN have extensive roles in the brain, revealed in studies using mouse models, including in synaptic transmission and plasticity, structural stability, and memory formation (Fawcett, Oohashi and Pizzorusso, 2019).

Importantly, CSPGs are up-regulated following neuronal injury and block myelin repair through oligodendrocyte inhibition (Lau *et al.*, 2013; Harlow and Macklin, 2014). When CSPGs are removed, this stimulates oligodendrocyte proliferation, making CSPGs a

potential target for repair after demyelination (Lau *et al.*, 2012). CSPGs distribution is also differentially affected during multiple sclerosis stages (Ghorbani and Yong, 2021). Versican and brevican are also abundant in developing white matter tracts (Ogawa *et al.*, 2001).

Another important role of CSPGs, in particular versican and brevican, is in inhibiting neurite outgrowth in the mature brain (Yamada *et al.*, 1997; Niederöst *et al.*, 1999; Schmalfeldt *et al.*, 2000), blocking regeneration of neurites following injury (Snow, Mullins and Hynds, 2001; Beller and Snow, 2014).

CSPGs are also accumulated at nodes of Ranvier, the gaps between myelinated sections along the axon, where they form the perinodal ECM. The perinodal ECM has roles in the formation and stabilisation of the nodes of Ranvier and in axon conduction (Fawcett, Oohashi and Pizzorusso, 2019). Its morphology is also altered in the absence of the CSPGs neurocan, versican or brevican (Bekku *et al.*, 2009; Dours-Zimmermann *et al.*, 2009; Bekku and Oohashi, 2011). CSPG expression varies across brain regions (Dauth *et al.*, 2016).

2.5.1. Versican

Versican has four isoforms resulting from alternative splicing: V0, V1, V2 and V3. V0, V1 and V3 are expressed in all tissue types, while V2 is solely expressed in the brain, predominating around myelinated fibres and at the nodes of Ranvier (Song and Dityatev, 2018). They have distinct developmental expression patterns (Milev *et al.*, 1998). The expression of V1 in the brain is known to be limited and confined to the choroid plexus, meninges, and the lining of blood vessels. V1 expression changes during development, with peak expression at birth. Functionally, in the brain, versican V1 is involved in the inhibition of oligodendrocyte precursor cells and remyelination (Ghorbani *et al.*, 2022). Versican V2 is essential for the formation of the extracellular matrix at the nodes of Ranvier (Dours-Zimmermann et al., 2009) and is the predominant versican isoform in the mature brain. The V0 isoform is very briefly

expressed during development (Perissinotto *et al.*, 2000). The V3 isoform is not as researched as its expression levels are also quite low (Harten *et al.*, 2020). Versican isoforms accumulate in diseased brain and spinal cord tissue in conditions such as cancer (Stephenson and Yong, 2018; Wight *et al.*, 2020) or in brain lesions (Zheng *et al.*, 2004; de Jong *et al.*, 2020).

2.5.2. Brevican

Alternative splicing also yields two brevican isoforms, one that is anchored to the cell membrane, and another more abundant one found throughout the ECM (Seidenbecher *et al.*, 1995). Brevican is found exclusively in the brain and its expression plateaus once adulthood is reached (Milev *et al.*, 1998). It is also located in perineuronal nets and diffusely in the white matter (Dours-Zimmermann et al., 2009).

Brevican located in PNNs controls the assembly of molecules responsible for synaptic transmission, affecting neuron excitability, the ability to respond to stimuli. Brevican levels are regulated by neuronal activity and brevican is required for learning and memory (Favuzzi *et al.*, 2017). Brevican shares similar functional roles with versican: it is present at the nodes of Ranvier and its absence leads to a reorganisation of the nodal ECM. The lack of brevican at the nodes of Ranvier can be compensated by the presence of versican V2 (Bekku *et al.*, 2009).

2.5.3. Imaging the extracellular matrix

Quantitative imaging of the extracellular matrix can help to further investigate its function and interactions with white matter. Optical, fluorescence, and electron microscopy are very useful approaches; however, they are invasive and usually focus only on small sections of tissue. MRI, particularly diffusion-weighted MRI and magnetic resonance spectroscopy (MRS) can provide complementary, whole-brain, non-invasive insight into yet unresolved biological mechanisms occurring at the level of the ECM (Soria *et al.*, 2020). However, these methods have limited specificity.

Chapter 3: Imaging principles: Magnetic resonance imaging and histology

Context

This chapter will introduce the principles of the imaging techniques used in this thesis – MRI and histology – starting from data acquisition to image processing considerations, with emphasis on rodent neuroimaging.

The first half of the chapter will present fundamental magnetic resonance imaging (MRI) theory and expand on structural and *ex-vivo* MRI. It will introduce and describe MRI methods that probe tissue microstructure, specifically diffusion-weighted MRI, relaxometry (focusing on R_2^* mapping) and quantitative susceptibility mapping (QSM). It will highlight the theory behind the signal modelling techniques used in this thesis, such as diffusion tensor imaging (DTI), neurite orientation dispersion and density imaging (NODDI) or QSM. Finally, it will introduce the field of imaging genetics and present the work that motivated this thesis.

The second half of the chapter will introduce the basics of histology, concentrating mostly on immunohistochemistry of formalin-fixed, paraffin-embedded samples – the technique used in this thesis. Finally, the chapter will conclude by presenting methods employed for MRI-histology data fusion and joint analysis, detailing those used in this work.

The information presented in this chapter is based on the following text books: *Principles of Magnetic Resonance Imaging,* by Dwight Nishimura (Nishimura, 1996), *Diffusion MRI: From Quantitative Measurement to In-Vivo Neuroanatomy,* edited by Heidi Johansen-Berg and Tim Behrens (Johansen-Berg and Behrens, 2009), Quantitative Magnetic Resonance Imaging, edited by Nicole Seiberlich, Vikas Gulani and others (Seiberlich *et al.*, 2020) and *Bancroft's Theory and Practice of Histological*

Techniques edited by Kim S Suvarna, Christopher Layton and John D Bancroft (Suvarna, Layton and Bancroft, 2013).

3.1. Fundamentals of magnetic resonance imaging

MRI has revolutionised both medicine and neuroscience, since Paul Lauterbur's first demonstration of its ability to produce images of the human body over 50 years ago (Lauterbur, 1973). It allows researchers to gain non-invasive insight into the structure and function of the whole brain using non-ionising radiation. MRI enables data acquisition at high resolution and in reasonable scan times. Nowadays, studies regularly acquire data with up to 0.1mm isotropic resolution in rodents in under 30 minutes. Although magnetic resonance is governed by the principles of quantum mechanics, a classical approximation is sufficient for its description and understanding (Hanson, 2008).

3.1.1. Magnetic resonance

MRI is based on the principle of magnetic resonance, which describes the interaction of atoms which have a quantum property known as "spin" with magnetic fields. Spin is an intrinsic type of angular momentum. Hydrogen nuclei (protons) in water molecules have a spin, meaning that, in a large magnetic field B_0, their net magnetisation M_0 aligns with B_0 and the magnetic moments of individual protons precess around B_0.

The amplitude of M_0 is proportional to the density of protons in that medium. The proton density varies with tissue type. B_0 establishes a coordinate system with the longitudinal component of M_0 aligned with B_0 (M_z, in the "z" direction) and the transverse component perpendicular to B_0 (M_{xy}, found in the "x-y" plane). Additionally, the net transverse component of M_0 is zero, due to the random orientation of the spin population in the transverse plane.

In a B_0 magnetic field, the atoms' magnetic moments will precess at a well-defined frequency, the Larmor frequency $\omega = ({}^{d\theta}/_{dt})$, according to:

$$\omega = \gamma B_0.$$

(3.1)

Here, γ is the gyromagnetic ratio of the nucleus, θ is the phase. For hydrogen atoms, γ is 42.57 rad · s^{-1} ·$Tesla^{-1}$. The B_0 field strengths of MRI scanners range from 1.5-21T, with the maximum field strength limited primarily by the size of the imaged object.

A radiofrequency (RF) magnetic pulse at the Larmor frequency, B_1, is produced by transmit coils and used to excite the precessing spins out of the equilibrium state (where M_0 is aligned with B_0). B_1 is perpendicular to B_0 and of magnitude much smaller than B_0 (<25μT) and short duration (~0.1-10ms). B_1 tilts the net magnetisation vector M_0 towards the transverse (x-y) plane, by applying a torque rotating M_0 by an angle (also known as "flip angle") dependent on B_1's magnitude and duration. The resulting M_{xy} component of M_0 is the detected signal.

A 90° RF pulse tips M_0 completely into the transverse (x-y) plane, where it continues precessing at the Larmor frequency, with $M_{xy} = M_0$. In the same way that the time-varying electric current in the coils produces a time-varying magnetic field RF pulse B_1, the rotating transverse magnetisation induces an electromotive force, driving an electric current in an RF receiver coil, according to Faraday's law. This recorded current provides the MR signal used to measure the size of the transverse magnetisation, M_{xy}. The same RF coil can fulfil both transmit and receive functions, or one can use separate coils for each function.

3.1.2. MR relaxation times

Following the application of the B_1 RF excitation pulse, the magnetisation will relax back into alignment with the B_0 field via two mechanisms, longitudinal and transverse relaxation, which occur simultaneously and independently. T_1 is the time constant characterising the longitudinal relaxation (**Figure 3.1** a), while T_2 characterises the

transverse relaxation (**Figure 3.1** b). Longitudinal relaxation is also known as "spin-lattice" relaxation, as it is due to interactions of the spins with their environment – the 'lattice'. By releasing energy into the external environment, the spins return to their energetically-favourable pre-excitation equilibrium state. Transverse relaxation is known as spin-lattice relaxation, being caused by the interactions and subsequent energy transfers between the neighbouring spins. These interactions lead to dephasing and a decay of the transverse magnetisation M_{xy}.

T_1 and T_2 values vary with tissue type and scanner field strength; for instance, *in-vivo* human brain tissue at 3T has grey matter T_1 values of about 1300-1400ms, whereas T_2 grey matter values are in the range of 100-130ms (Wansapura *et al.*, 1999). Typically, cerebrospinal fluid (CSF) has the largest T_1 and T_2 values, followed by grey and then white matter.

Following excitation, the longitudinal magnetisation M_z returns to alignment with the B_0 field with an exponential timescale T_1 described as:

$$M_z(t) = M_0\left(1 - e^{-t/T_1}\right). \tag{3.2}$$

The transverse component M_{xy} of the magnetisation decays exponentially as the spins dephase. This occurs with time constant T_2 after the 90° pulse:

$$M_{xy}(t) = M_0 e^{-t/T_2}. \tag{3.3}$$

B_0 inhomogeneities due to the experimental setup or magnetic susceptibility variations (e.g., at the air-tissue interface) lead to additional spin dephasing and a faster M_{xy} decay, with a time constant $T_2^* < T_2$, where

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}.$$

(3.4)

These magnetic susceptibility variations can provide us with further insight into microstructural properties of tissue. If one is not interested in the effect of inhomogeneous susceptibility, this additional signal decay can be reversed by applying a 180° "refocusing RF pulse".

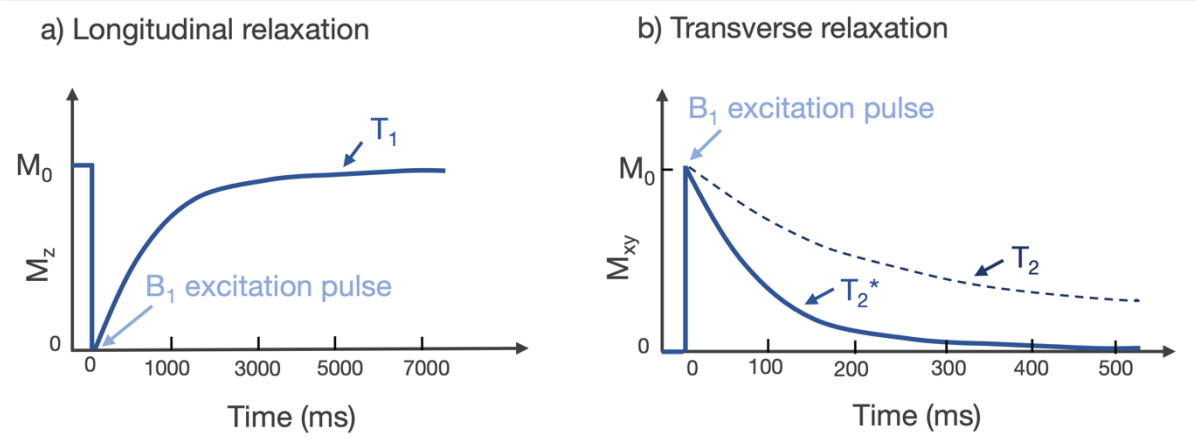

Figure 3.1. a) Longitudinal relaxation: After the B_1 RF excitation pulse is applied (time=0), the longitudinal component of M_0 (M_z, oriented in the z direction) returns to equilibrium with the time constant T_1, the longitudinal relaxation time. b) Transverse relaxation: After the B_1 RF excitation pulse is applied (time=0), the transverse component of M_0 (M_{xy}, in the x-y plane) decays with time constant T_2 (in a spin-echo experiment) and T_2^* (in a gradient-echo experiment).

The timings of RF pulses and tissue relaxation times (T_1, T_2, T_2^*) impact MR contrast, the signal difference between the tissues of interest. MRI acquisitions are modular, containing an image formation component and a contrast component, which can be combined depending on the experimental aims. In the following sections, we will introduce these concepts and their relevance to this work.

3.1.3. MR image formation

The principles presented so far enable signal detection and contrast. This signal represents the bulk response of the whole spin population within the object placed in the MRI scanner. This bulk response can be manipulated to yield further information. Magnetic field gradients are used to achieve spatial localisation of the magnetic response and obtain an image. Three orthogonal magnetic gradients can be produced by gradient coils. The field generated by the coils is oriented in the same direction as the B_0 field, but its strength varies linearly along a specific direction. For instance, we can define the gradient field strength in the x-direction, as

$$G_x = \frac{dB_z}{dx}. \tag{3.5}$$

Gradient strength is measured in mT/m, and in a clinical scanner it is on the order of 50mT/m. If three perpendicular gradients are applied at any location (x, y, z), the total field at that location, oriented in the direction of the B_0 field, is:

$$B_z(x, y, z) = B_0 + \frac{dB_z}{dx}x + \frac{dB_z}{dy}y + \frac{dB_z}{dz}z. \tag{3.6.}$$

From the Larmor equation, it results that the Larmor frequency at any position (x, y, z) is now of the form:

$$\omega(x, y, z) = \gamma(B_0 + G_x x + G_y y + G_z z). \tag{3.7}$$

Image formation generally involves different kinds of spatial localisation: slice/slab selection, frequency (or "readout") encoding and phase encoding, all of which use the scanner's gradient coils. In general, there are two methods used in MRI for encoding spatial information. 2D imaging first uses RF pulses alongside gradients to excite the magnetisation from thin planes of the object ("slice selection"). The slice-selection gradient is applied in a direction perpendicular to the desired imaging plane simultaneously with a B_1 RF excitation pulse, ensuring that only the spins within the frequency bandwidth of the RF pulse are excited and will contribute to the final image. Alternatively, 3D imaging excites the entire imaging volume (or thicker slabs) and then manipulates the gradients over multiple measurements so that the signal can be localised within a 3D volume.

The image-encoding gradients – frequency and phase encoding – are then applied in the two perpendicular directions within the excited slice, after the RF excitation pulse. This results in a spatially modulated phase of the precessing spins in the x-y plane. This phase also depends on the integrated area of the gradients, $\theta = 2\pi k$, where:

$$\boldsymbol{k} = \gamma \int_0^t \boldsymbol{G}(\tau) d\tau. \tag{3.8}$$

The complex signal measured by the receive coil is the integral of the signal from all the spins multiplied by the spatial frequency variable $\boldsymbol{k}$. For instance, in our 2D example:

$$S(\boldsymbol{k}) = \int s\,(\boldsymbol{r}) e^{i\boldsymbol{k}\boldsymbol{r}} d\boldsymbol{r}.$$

(3.9)

Here, $s(\boldsymbol{r})$ is the signal at a given point $\boldsymbol{r}$. $\boldsymbol{k}$ defines the position in k-space (in the 2D case, it is denoted by k_x and k_y). The domain in which data is acquired is known as the 'k-space'. As we have a 2D k-space, $S(\boldsymbol{k})$ at each (x, y) data point represents the signal contribution at each spatial frequency. By taking the 2D Fourier transform of the signal measured at each location in k-space, we can reconstruct an image of the signal at each point in the object. Equivalently, 3D image reconstructions uses 3D Fourier transformation. The centre of k-space measures the mean signal intensity across the image. Values near the centre of k-space encode lower spatial frequencies. The periphery of k-space encodes higher spatial frequencies, such as sharp boundaries.

k-space is sampled discreetly and finitely, and is related to image characteristics as follows:

$$\Delta k = \frac{1}{FOV}.$$

(3.10)

Δk is the distance between two points in k-space and FOV is the size of the field of view. Similarly,

$$\Delta x = \frac{1}{2|\boldsymbol{k}_{\max}|}.$$

(3.11)

Here, Δx is the image resolution and k_{max} is the extent of the sampled k-space, which is limited by the experimental parameters.

k-space sampling strategies – the trajectories through which information is collected in the 2D/3D k-space domain – are dictated by the sequence of gradients. Often, a cartesian trajectory is used, where lines of k-space are filled one at a time (in the k_x direction). The readout/frequency-encoding gradient G_x ensures the coverage of the k-space in the k_x direction. The phase-encoding gradient G_y values are changed incrementally between excitations to ensure the next line of k-space is reached before the G_x gradient is applied. The height of the two gradients represents their strength; together with their duration they determine how far in the k-space one travels prior to data acquisition.

3.1.4. Pulse sequences

MRI measurements require that we manipulate RF pulses and gradient fields, alongside the timings of signal recordings, to produce images reflecting the tissue properties of interest. The precise timing of these events is referred to as a "pulse sequence" and there are different combinations of RF pulses and gradient sequences that produce different types of MRI images – also known as "MRI modalities". These MRI modalities provide different kinds of information about the tissue of interest. Within a pulse sequence, a slice-selective RF pulse is applied, followed by gradients in two perpendicular directions within the slice. The time between the RF excitation pulse and the centre of the echo is the echo time (TE). This occurs half-way during the rephasing lobe of the frequency-encoding gradient. The time between successive pulse sequences is the repetition time (TR).

An example pulse sequence is the 2D Fourier Transform Gradient-Echo (2DFT GRE) pulse sequence (**Figure 3.2** a). GRE sequences contain an excitation pulse followed by a readout. GRE sequences do not remove the effects of magnetic field inhomogeneities, meaning that the transverse magnetisation decays quicker, with the T_2^* time constant (**Figure 3.1** b).

An alternative approach used to reduce acquisition time considerably, echo-planar imaging (EPI), is used to traverse k-space in a single TR (**Figure 3.2** b). Here, the application of intermittent ("blipped") phase-encoding gradients allows for the k-space to be traversed in a single line during one TR. In this thesis, a 3D multi-echo gradient-echo sequence is used to acquire *ex-vivo* MRI data. This data is then used for T_2^*- and quantitative susceptibility mapping (QSM). Specific details of this this protocol will be covered in **Chapter 4**.

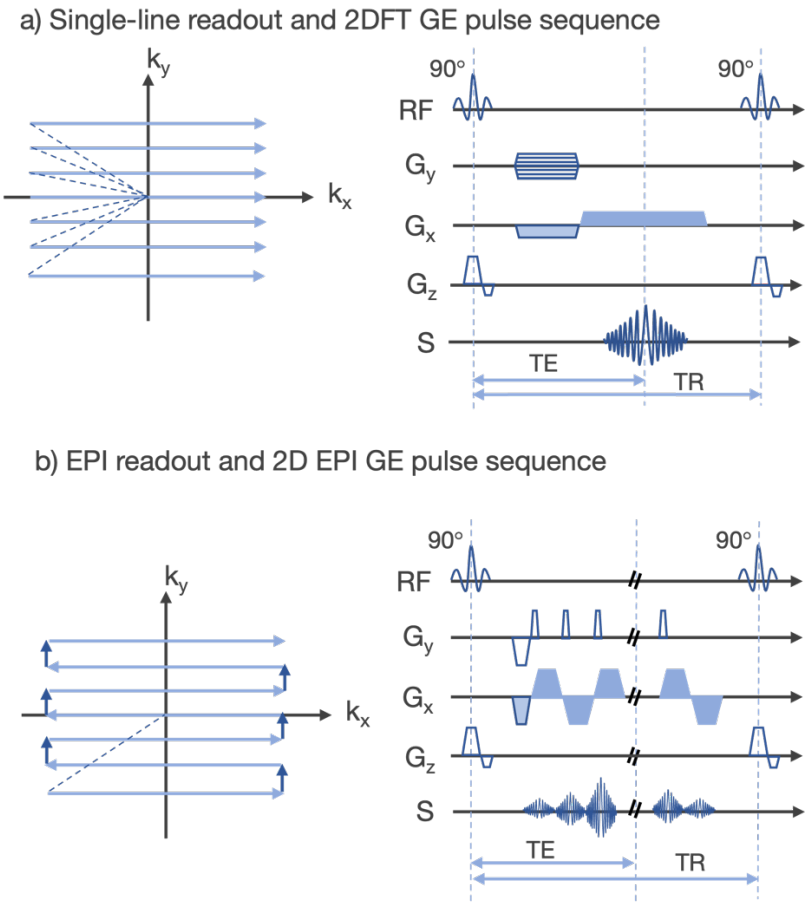

Figure 3.2. The gradient echo sequence trajectories through k-space and pulse sequence for either a) single-line or b) EPI acquisition. S is the signal, G_x is the readout gradient, G_y is the phase-encoding gradient and G_z is the slice-selective gradient.

Spin-echo (SE) sequences are a different kind of pulse sequences that specifically avoid the signal loss due to T_2^*. Like GRE sequences, they can be used in combination with various image formation schemes (e.g., EPI) and acquire signal at time TE. Distinctively, however, SE sequences recover the signal loss due to magnetic field inhomogeneities by applying a 180° refocusing pulse before the readout gradient. Their signal levels are thus dictated by the T_2 time constant.

In this thesis, structural MRI data is acquired using the Turbo Rapid Acquisition with Relaxation Enhancement (TurboRARE) variant of the spin-echo sequence, equivalent

to the turbo spin-echo (TSE) sequence. In this sequence, multiple spin-echoes are generated using 180° pulses and phase-encoded separately, accelerating the acquisition. We use this structural MRI data to examine volumetric differences across the mouse brain and for the MRI-histology registration. The details of our protocol will be covered in **Chapter 4**. The MRI-histology registration will be covered in **Chapter 5**.

3.1.5. Image contrast

MR image contrast depends on tissue composition (reflected by the proton density and the relaxation times T_1, T_2 and T_2^*), on the degree of diffusion weighting (which will be described below), as well as on the manipulation of the timings of RF pulses.

The signal (S, equivalent to M_{xy}) equation form at time TE is as follows:

$$S = S_0 \frac{sin\alpha(1 - e^{-TR/T_1})}{(1 - e^{-TR/T_1 cos\alpha})} e^{-TE/x}. \quad (3.12)$$

For spin-echo sequences, x is T_2 , for GRE sequences, x is T_2^*. α is the flip angle of the refocusing pulse.

A very short TE ($TE \ll T_2$) and a moderate TR ($TR < T_1$) cancel out the T_2 contributions to the signal, resulting in an image in which the signal contrast is dominated by the T_1 of tissue ("T_1-weighted"). Tissue with shorter T_1 appears brighter on T_1-weighted images: for instance, on brain MRI, white matter appears brighter than grey matter and a good grey-white matter contrast can be observed. Likewise, a long TE (close to T_2) and long TR ($TR \gg T_1$) results in a T_2-weighted image with T_2 contrast. In T_2-weighted images, tissue with longer T_2 values looks brighter. Cerebrospinal fluid has the longest T_2 values in the human brain and appears extremely bright on T_2-weighted

images. Data acquired using SE sequences can be either T_1- or T_2-weighted. GRE data can likewise be T_1- or T_2^*-weighted.

3.1.5.1. Diffusion contrast

A different kind of contrast, diffusion contrast, is achieved using diffusion-weighted MRI (dMRI). dMRI uses diffusion-weighting gradients – by activating the same gradient coils used for spatial encoding – to collect microstructural information from the tissue. This is done by sensitising the images to the diffusion of water molecules and measuring their voxel-averaged random motion in tissue. This motion is influenced by the tissue microarchitecture.

A basic diffusion-weighted spin-echo sequence (Tanner and Stejskal, 1968), involves the addition of diffusion-weighting gradients (**Figure 3.3**). The first gradient creates a phase modulation across the spins, while the second gradient reverses this effect. If the spins are stationary, then the application of the two gradients will have no effect and the net change in overall phase of these spins will be zero. However, if the spins are diffusing along the direction of the diffusion gradient, ending up in a different location during the application of the second diffusion gradient, then the net phase difference accrued will be proportional to their displacement in that direction. This phase difference effectively results in more dephasing along the direction of the gradient pair, thus further reducing the net magnetisation and resulting signal.

In this thesis, we utilise a spin-echo segmented EPI diffusion-weighted sequence for the acquisition of dMRI data. The parameters of this sequence will be presented in **Chapter 4**. Segmented EPI data only acquires a fraction of k-space during each TR. This mitigates distortion artifacts which are usually present when acquiring all data using a single-shot EPI sequence.

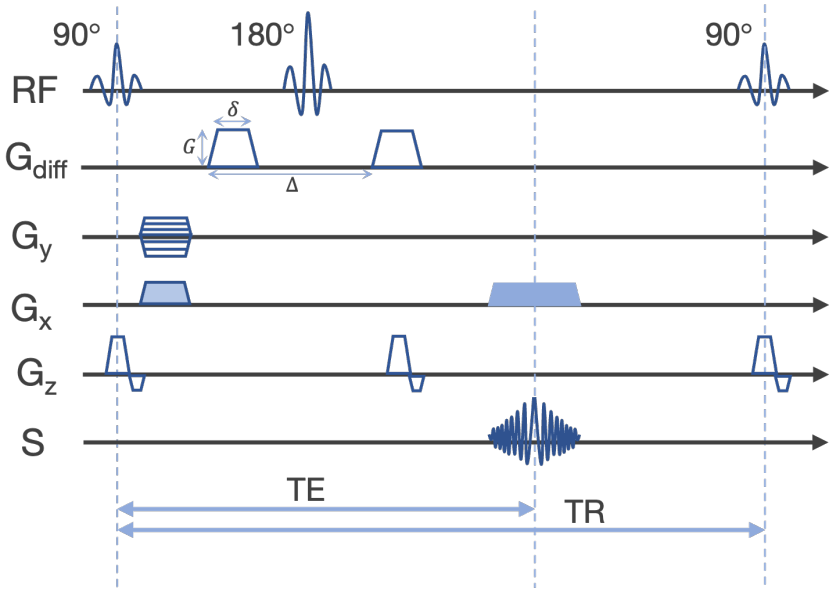

Figure 3.3. Spin-echo diffusion-weighted sequence. G is the amplitude of the gradient, δ is its duration and Δ is the time interval between the start of the first gradient and the start of the second gradient.

dMRI data exploits the diffusion of water molecules to generate diffusion contrast. The diffusion contrast is influenced by microscopic tissue features such as cellular boundaries, shapes and sizes. The key parameter of the diffusion sequence, the b-value b, describes the amount of diffusion-weighting, where:

$$b = \gamma^2 |\boldsymbol{G}|^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right). \tag{3.13}$$

The parameters G, δ and Δ (**Figure 3.3**) relate to the timings and amplitude of the diffusion gradients. In the case of a spin-echo sequence, the diffusion-weighted signal equation at TE is:

$$S(b, D) = S_0 \left(1 - e^{-TR/T_1}\right) e^{-TE/T_2} e^{-bD}. \tag{3.14}$$

Here, D is the apparent diffusion coefficient (ADC), the diffusion coefficient of water molecules in the presence of boundaries. Within brain tissue, the diffusion of water molecules is affected by the presence of cell membranes, macromolecules, or intracellular structures. In addition, water diffuses more freely along the direction of white matter fibres than across them. This has the potential to yield information about tissue microstructure.

Tissue water ADC is lower than the diffusion coefficient of water molecules in a solution. The free diffusion coefficient, D_{free}, depends on the mean-squared displacement of the particles, x^2 , the number of dimensions considered (N, where $N = 3$ in the case of a volumetric measurement) and the diffusion time t (Einstein, 1905),

$$D_{free} = \frac{x^2}{2Nt}. \tag{3.15}$$

The diffusion coefficient is characteristic of every medium or tissue type and depends on the size of the diffusing molecules and the temperature and viscosity of the environment, among other factors. The diffusion coefficient of water molecules in solution at 25°C is approximately 3μm^2/ms. In a boundary-free 3D medium, the probability distribution of the particle displacement is described by an 3D Gaussian (normal) distribution.

The phase incoherence exhibited by diffusing spins during diffusion-weighting, and hence the signal attenuation and image contrast, are dictated by the strength and duration of the applied gradients. Diffusion-weighting gradients achieve G values of up to 300mT/m (Huang *et al.*, 2015) and preclinical scanners going up to 2500mT/m (Johnson *et al.*, 2023).

dMRI data is usually acquired by repeating measurements for multiple orientations along a sphere ("a shell") to gather information about water diffusion in all three dimensions. When these measurements are acquired at the same b-value, where the b-value dictates the radius of the sphere, the whole dataset is said to be acquired at the same shell. We can also acquire data at multiple orientations and multiple b-values, known as "multi-shell acquisition", which allows for the characterisation of water diffusion at different levels of diffusion-weighting. Mathematical modelling of the dMRI signal uses either single-shell or multi-shell acquisitions to produce estimates of tissue microstructural parameters, such as tract orientation. The number of gradient directions and their distribution within a dMRI dataset determines the angular resolution. Data without any diffusion weighting (b=0) is usually also acquired, to measure S_0 (**Equation 3.13**).

3.1.5.2. Contrast enhancement

The contrast between different tissue components can be further enhanced using contrast agents. Gadolinium-based (Gd-based) contrast agents act by shortening the relaxation times of the hydrogen protons, primarily T_1. In *ex-vivo* tissue, for example, T_1 drops down from about 500ms to about 100ms in grey matter at 4mM Gd-based concentration, whereas T_2 gets reduced from about 70ms to 30ms in grey matter (4.7 T field strength) (D'Arceuil, Westmoreland and de Crespigny, 2007). In clinical settings, contrast agents can be administered orally or intravenously. In *ex-vivo* preclinical work, researchers perfuse the animals with a solution containing a Gd-based contrast agent and fixative, which ensures uniform delivery of the contrast agent (Huang *et al.*, 2009).

3.1.6. Signal-to-noise ratio

Signal-to-noise ratio (SNR) is a central factor in the optimisation of MRI acquisition protocols. It is defined as the mean signal divided by the standard deviation of the

noise. It is influenced by B_0 field strength, voxel volume, receive coil sensitivity profile, coil diameter, number of signal averages and bandwidth. Preclinical MRI systems have higher field strengths and smaller coils, producing high-SNR and high-resolution images highlighting exquisite tissue detail. While in clinical systems the main noise source is thermal noise from the sample, the top noise contributor on preclinical systems is electronic noise from the RF coils. To reduce this noise, state-of-the-art preclinical systems come equipped with cryogenically-cooled RF coils.

3.1.7. Structural MRI processing and analysis

High-resolution, high-SNR MRI can provide detailed anatomical information. When analysing structural MRI data, we can exploit the differences in contrast between different areas of tissue, such white versus grey matter or cortical versus deep grey matter areas.

Segmentation of brain areas and the extraction of volumetric information in mice usually requires the registration (alignment) of the brains to a template mouse brain and associated labelled atlas. Registration seeks to find a transformation mapping between two images that achieves maximum similarity. One very popular rodent brain registration and atlas-based segmentation pipeline (Chakravarty *et al.*, 2013; Friedel *et al.*, 2014), also used in this thesis, uses linear and non-linear stages to estimate the non-linear displacement fields that provide a measure of local volume expansion and contraction (Lerch, Sled and Henkelman, 2010), alongside an atlas-based segmentation.

3.1.8. *Ex-vivo* MRI: Imaging considerations

Ex-vivo MRI provides a link between *in-vivo* MRI and histology. *Ex-vivo* rodent tissue is usually obtained via transcardiac perfusion fixation, a process which involves delivery of fixative and contrast agent (e.g., Gd-based) via vascular injection while the

animals are under anaesthesia. The samples are then stored at low temperatures for long periods. Tissue macro- and microstructure are generally preserved following correct fixation, with little but consistent tissue shrinkage (around 5%). To ensure that this shrinkage has stabilised across samples, it is recommended that they are scanned after the first month of storage (de Guzman *et al.*, 2016). To limit brain shrinkage and significant SNR reduction, samples should not be stored in fixative long-term (de Guzman *et al.*, 2016). The proton density and relaxation times of *ex-vivo* samples are reduced compared to *in-vivo* (Yong-Hing *et al.*, 2005; Roebroeck, Miller and Aggarwal, 2019). Additionally, they show a dependence on Gd-based contrast agent concentration (D'Arceuil, Westmoreland and de Crespigny, 2007). Provided the optimal Gd-based contrast agent concentration is chosen, the MR image contrast should be equivalent to *in-vivo* scenarios.

Further, due to the presence of the fixative, the diffusivity is significantly reduced (Sun *et al.*, 2005; Dyrby *et al.*, 2011). Diffusivity is also affected by the temperature. Other tissue parameters, such as magnetic susceptibility or diffusion anisotropy, are not known to be affected by fixative (Sun *et al.*, 2005; Liu *et al.*, 2011).

In the *ex-vivo* scenario, due to decreased diffusivity, a higher b-value is needed to achieve equivalent diffusion contrast to *in-vivo*. Studies have shown that *ex-vivo* b-values of 4000s/mm^2 provide similar signal attenuation to *in-vivo* b-values of 1000s/mm^2 (D'Arceuil, Westmoreland and de Crespigny, 2007).

3.2. Imaging microstructure using quantitative MRI

In addition to the bulk anatomical insight provided by structural MRI, microstructural MRI helps us understand the molecular drivers of tissue change. dMRI, relaxometry and quantitative susceptibility mapping (QSM) fall under the hood of microstructural MRI and can each grant complementary insight into tissue properties.

3.2.1. Diffusion-weighted MRI

3.2.1.1. Data processing

dMRI datasets require processing for artifact removal prior to being fitted using mathematical models. These processing pipelines minimise image artifacts and facilitate signal modelling and data analysis.

While subject motion is a major source of dMRI artifacts *in-vivo*, this is not an issue *ex-vivo*. Here, image distortion due to eddy currents can be modelled and corrected (Andersson and Sotiropoulos, 2016; Ma *et al.*, 2020). Susceptibility distortion due to B_0 inhomogeneities manifests as geometric warps and signal pile-up. This off-resonance, susceptibility-induced field can be modelled, for instance by using two acquisitions using reverse phase-encoding directions, and subsequently corrected, like FSL's TOPUP does (Andersson, Skare and Ashburner, 2003). Gibbs-ringing, visible at sharp tissue boundaries, can also be eliminated (Kellner *et al.*, 2016). Signal drift resulting from scanner instabilities (e.g., heating) can be modelled using interleaved b=0 volumes (Hansen *et al.*, 2019). Other important factors that are often tackled are thermal noise, Rician bias or gradient non-linearities (Schilling et al., 2022; Tax et al., 2022). These issues, as well as important considerations such as managing vendor-specific data structures and formats on a preclinical system (7T Bruker BioSpec 70/20 USR) will be rigorously examined in **Chapter 4**, which will present the development of our *ex-vivo* rodent brain dMRI processing pipeline.

3.2.1.2. Data modelling

A broad range of signal representations and biophysical models are used to characterise the dMRI signal. The work in this thesis is focused solely on the analysis of white matter using the diffusion tensor imaging (DTI) signal representation (Basser, Mattiello and LeBihan, 1994) and the neurite orientation dispersion and density imaging (NODDI) model (Zhang *et al.*, 2012).

3.2.1.2.1. Diffusion tensor imaging

DTI is a description of the anisotropic free diffusion of water molecules using a 3D Gaussian model. The coefficients of this rank-2 tensor, estimated via linear regression of the log-transformed signal, describe Gaussian diffusion with different diffusivities along three eigenvectors. The average of the tensor's eigenvalues (λ_1, λ_2, λ_3) is known as the mean diffusivity (MD):

$$MD = \lambda = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \tag{3.16}$$

The fractional anisotropy (FA) describes how strongly anisotropic the diffusion profile is:

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \lambda)^2 + (\lambda_2 - \lambda)^2 + (\lambda_3 - \lambda)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{3.17}$$

FA ranges from 0 – corresponding to isotropic diffusion – to 1 – corresponding to diffusion constrained to a single axis. The eigenvectors of the tensors provide information regarding the orientation of the tensor. The principal eigenvector describes the principal direction of diffusion; in white matter, its orientation provides information regarding the major tract orientation.

The diffusion tensor can reflect relevant properties of tissue microstructure. The orientation of the principal eigenvector indicates the principal diffusion direction and

the approximate orientation of a given tract. For instance, low FA values are found in the CSF, while high FA usually corresponds to white matter voxels.

One of the main downsides of DTI is its lack of specificity, only indirectly characterising tissue. For example, a decrease in fractional anisotropy in a region can result from a range of cellular processes (e.g., demyelination, axonal damage or cell death) which cannot be disentangled in the absence of additional information (e.g., histology). Therefore, characterising the source of the change in parameters can be difficult. This has not stopped DTI from gaining incredible popularity in the community, with notable applications in the clinical domain (Sundgren *et al.*, 2004).

To estimate the coefficients of the diffusion tensor, at least 6 independent volumes acquired at different orientations are required, as well as a b=0 volume. This acquisition is done at the same shell. In practice, however, to mitigate the effects of noise and obtain a better fit, 20-30 directions should be obtained (Jones, 2004).

3.2.1.2.2. The NODDI model

Contrastingly to DTI, which is a signal representation that can be applied across all tissue types, the NODDI model is a biophysical model that describes the diffusion of water molecules in white matter. NODDI makes inferences regarding tissue microstructure, with the goal of overcoming DTI's lack of specificity. NODDI models three compartments: intra-axonal ($intra$, neurite or intra-cellular volume/water fraction, ICVF), extra-cellular ($extra$) and isotropic (iso, or CSF) compartment. The normalised signal A is represented as:

$$A = (1 - f_{iso})(f_{intra}A_{intra} + (1 - f_{intra})A_{extra}) + f_{iso}A_{iso}. \quad (3.18)$$

NODDI fitting requires data at two different b-values. ICVF is modelled as a set of cylinders (sticks or fibres), with unhindered diffusion along the length of the axons and no diffusion across the axons. The sticks can be either parallel or highly dispersed, with the fibre orientation distribution (FOD) modelled as a Watson distribution. The intra-axonal signal compartment is modelled as the convolution of the Watson FOD with a fibre-response function of the fibres, assuming a fixed parallel diffusivity (*in-vivo* this was set to 1.7μm^2/ms in the original paper). The extra-axonal space assumes hindered but unrestricted diffusion – i.e., Gaussian diffusion with reduced diffusivity – and is modelled as the convolution of a Watson FOD with an axially-symmetric diffusion tensor. The CSF compartment is modelled as isotropic Gaussian diffusion, with fixed isotropic diffusivity (*in-vivo*, this was set to 3μm^2/ms in the original paper). Based on the Watson distribution, one can derive an orientation dispersion index (ODI), reflecting tract disorganisation within a voxel and ranging from 0 (perfectly aligned fibres) to 1 (isotropic dispersion). A graphical representation of the NODDI model is presented in **Figure 3.4.**

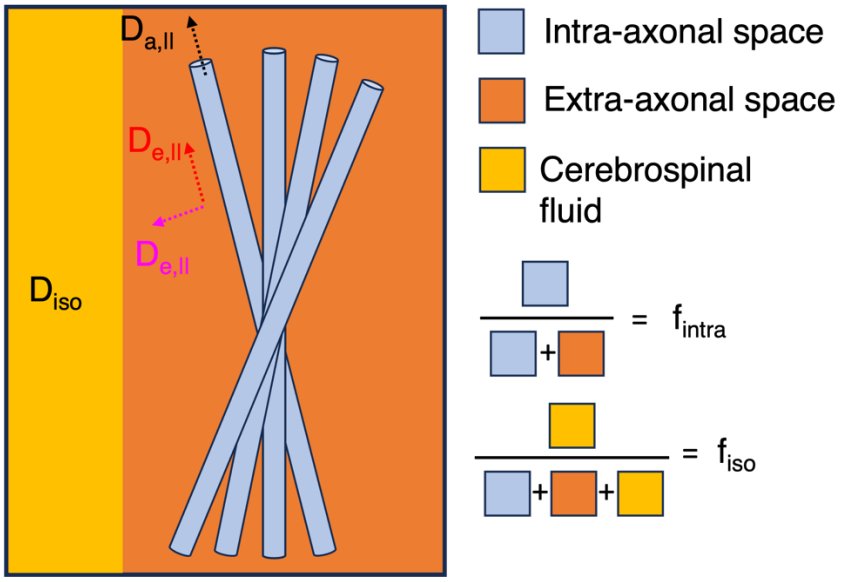

Figure 3.4. Graphical representation of the NODDI model adapted from Jelescu and Budde (Jelescu and Budde, 2017). $D_{a,||}$: parallel intra-cellular (axonal) diffusivity, $D_{a,\perp}$: perpendicular intra-cellular (axonal) diffusivity, $D_{e,||}$: parallel extra-cellular diffusivity, $D_{e,\perp}$: perpendicular extra-cellular diffusivity.

3.2.2. T_2^* Relaxometry: Data processing and fitting

MRI relaxometry refers to the voxel-wise mapping of relaxation times (T_1, T_2, T_2^*), with the aim of providing a quantitative characterisation of tissue. This section will describe T_2^* mapping, this being the technique used in this thesis.

T_2^* mapping relies on fitting **Equation 3.12** to data acquired using a gradient echo sequence. Bespoke sequences which reduce acquisition time, such as 2D/3D multi-echo gradient echo (GE) sequences (Denk and Rauscher, 2010), can be used. Multi-echo GRE sequences consist of an excitation pulse, succeeded by a series of refocusing pulses. The choice of TEs is important to make sure that the signal decay curve is uniformly covered (starting with the minimally achievable TE) and avoiding biasing estimates. Importantly, one needs to choose the longest TE carefully, ensuring it is long enough to capture the asymptote of the signal decay curve (**Figure 3.1**, b), yet short enough to minimise the scan time.

Data acquisition protocols for *ex-vivo* rodent T_2^* mapping should be standardised across cohorts, with temperature carefully monitored, as this can affect T_2^* estimates (He *et al.*, 2009). Prior to fitting, data can be subjected to minimal processing. This can include averaging of repeats, as well as Gibbs-ringing correction. A simple non-linear least square fitting algorithm, which also incorporates an additive noise component, can be used.

3.2.3. Quantitative susceptibility mapping: Data processing and fitting

The degree and polarity of magnetisation of any material following the application of an external magnetic field is known as magnetic susceptibility (usually denoted with χ). Diamagnetic materials show a negative response to a magnetic field, with a magnetic field opposed to the applied field. Paramagnetic materials have the opposite response, with a magnetic field within the sample that is in the direction of the applied magnetic field. In tissue, myelin and calcium are diamagnetic, whereas iron and deoxyhaemoglobin are paramagnetic.

Quantitative susceptibility mapping (QSM) measures the magnetic susceptibility of tissue using MRI. The same GRE sequence used for T_2^* mapping can be used. The estimation of χ necessitates data acquired at multiple echoes (a minimum of 2, but desirably more to acquire T_2^* estimates in the same dataset). In contrast to the modalities and models presented so far, QSM uses the phase of the signal. QSM is based on the principle that the phase component of the images is affected by the varying χ distribution in tissue. QSM images of the brain show exquisite contrast, particularly due to the diamagnetism exhibited by myelinated white matter and strong paramagnetism of iron-laden deep grey matter.

The phase has two components:

$$\theta = \theta_0 + \gamma \Delta B_z TE, \tag{3.19}$$

with θ_0 representing a coil-sensitivity dependent phase offset, γ the gyromagnetic ratio, ΔB_z the field variation in the z-direction, produced by all susceptibility sources within the imaged volume, and TE representing the echo time. In image space, the field variation is described as the convolution – denoted by the mathematical operator

$\otimes$ – between the susceptibility values χ and the unit dipole kernel d, which dictates how the object's magnetic susceptibility relates to variations in B_z:

$$\Delta B_z(\boldsymbol{r}) = B_0 d(\boldsymbol{r}) \otimes \chi(\boldsymbol{r}), \tag{3.20}$$

where $\boldsymbol{r}$ represents a given point.

QSM pipelines use complex data acquired using a multi-echo GRE sequence. Sometimes, coil combination is performed separately off-line. Some algorithms allow the removal of the phase offset θ_0 at the same time (Robinson *et al.*, 2017). A subsequent optional stage is registering the data (e.g., to the first volume) if there is misalignment between volumes. The magnitude data is used for this step. Next, the field map (ΔB_z, **Equation 3.20**) is estimated, by fitting the complex or the unwrapped phase data measured at multiple echo times. This is followed by background field removal in the ROI (e.g., the masked brain), to eliminate large field variations caused by air-tissue interfaces, which dominate the more subtle variations from local tissue susceptibility. Background field removal techniques include high pass or Laplacian filtering techniques (Schweser *et al.*, 2011), which assume that susceptibility sources from outside the region of interest, such as sinuses, are harmonic inside the region of interest and can hence easily be removed.

Finally, the susceptibility map is calculated using **equation 3.20**. **Equation 3.20** corresponds to a deconvolution of the dipole kernel from the ΔB_z map in the image domain, which is both computationally intensive and an ill-conditioned operation. Alternatively, this operation can be performed in the frequency domain, where it becomes a division by the Fourier transform of the dipole kernel. In this domain, the ill conditioned nature is manifested by division by zero values at the magic angle (54.7°) in k-space. There are various algorithms that have been developed (Haacke *et al.*, 2010; Li *et al.*, 2015; Wei *et al.*, 2015) to avoid this division by zero, all of which incur some degree of artifacts in the final χ map.

3.3. Imaging genetics

The versatility of MRI contrasts and its ability to provide anatomical, microstructural and functional information have enabled researchers to extract thousands of phenotypes from this type of data, such as the volume of specific brain structures or cortical thickness measurements. Imaging genetics is the area of research concerned with identifying and characterising genetic variants that are associated with IDPs (Ge, Schumann and Feng, 2013). Imaging genetics has yielded insights into diseases including neurodegenerative disorders (Xin *et al.*, 2022), psychiatric disease (Bogdan *et al.*, 2017) and brain development.

Until recently, the field of imaging genetics used small cohorts to investigate candidate genes known to be linked with certain diseases. The advent of population imaging studies like UK Biobank now enables larger scale "discovery" imaging genetics investigations where no candidate genes are specified a priori. The brain imaging component of the UK Biobank will ultimately include data from 100,000 participants. To capture a broad range of phenotypes, multi-modal MRI data was acquired, including structural (T_1- and T_2-weighted), functional (task and resting-state), T_2^*- and diffusion-weighted MRI.

A fully-automated processing pipeline (Alfaro-Almagro *et al.*, 2018) was designed to extract almost 4,000 IDPs per subject from this dataset. Acquisition and post-processing details for the UK Biobank protocol are provided in **Chapter 4**, in order to compare it against our *ex-vivo* multi-modal rodent brain MRI protocols, which were harmonised to UK Biobank.

Genome-wide association studies of brain imaging phenotypes in UK Biobank have yielded fruitful results. The first major research output (Elliott et al., 2018) used data acquired in the first 10,000 participants to reveal genetic loci (single nucleotide polymorphisms, or SNPs, presented in **Chapter 2**) associated with various IDPs spanning all MRI modalities. It included both novel findings as well as results which replicated loci identified by similar large-scale projects, such as ENIGMA (Hibar *et al.*,

2015). Many of these SNPs were either in linkage disequilibrium (LD) with SNPs in genes of interest or are eQTLs (a genetic locus influencing gene expression), which renders their interpretation difficult.

Previously-undescribed and noteworthy links include an association of a SNP located in an intron of the *VCAN* gene with multiple dMRI measures in most white matter tracts (199 IDPs). When using this SNP as a spatial regressor in dMRI data, the associated voxels span white matter tracts across the whole brain (**Figure 3.5**). The most common category of IDPs to associate with this SNP is mean NODDI ICVF parameter calculated in individual tracts, but there were also a limited number of hits for FA and MD IDPs. More recently, hits were identified between SNPs in *VCAN* and cortical grey-white matter contrast (Smith *et al.*, 2021). This IDP is calculated using the T_1-weighted structural data (Salat *et al.*, 2011; Mann *et al.*, 2018). 114/116 grey-white matter contrast IDPs showed statistically significant associations with *VCAN* SNPs, indicating brain-wide effects as well. Finally, statistically-significant associations between susceptibility measurements in white matter hyperintensities and SNPs in *VCAN* were also reported recently (Wang *et al.*, 2021).

Another intriguing result was that of associations between a SNP in an eQTL of *BCAN* with the averaged ICVF in the genu of the corpus callosum.

As detailed in **Chapter 2**, *VCAN* and *BCAN* encode extracellular matrix proteins known to inhibit myelin and neurite outgrowth. Their comparable roles and links to similar neuroimaging phenotypes suggest a common underlying biological mechanism leading to the SNP-IDP associations in both genes.

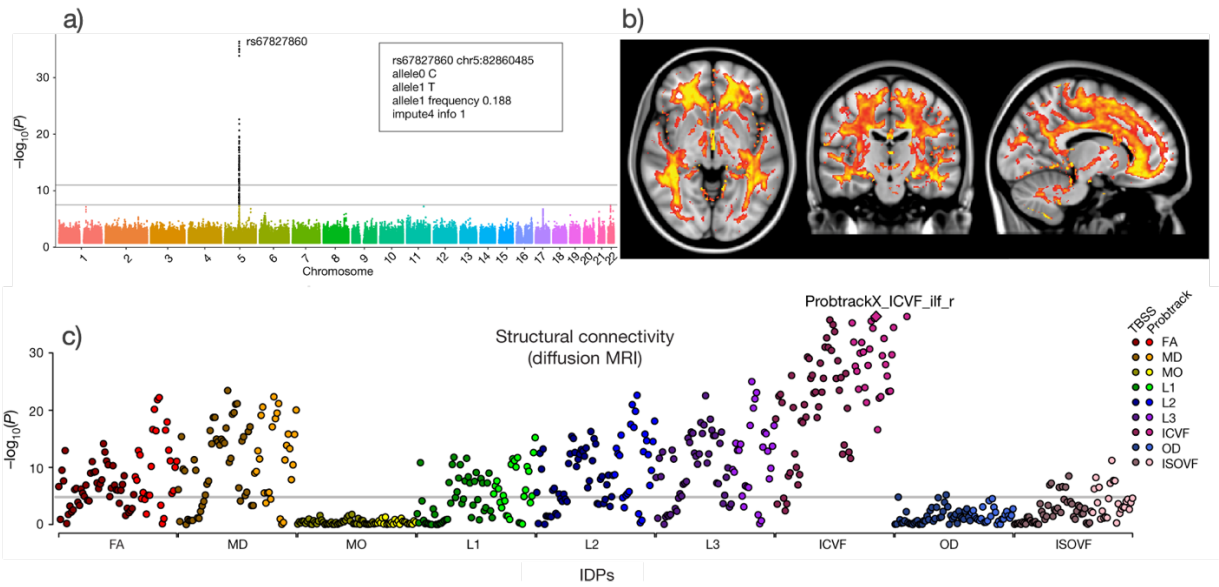

Figure 3.5. a) Manhattan plot of -$\log_{10}$(p-values) relating the ICVF measure with the strongest association (right inferior longitudinal fasciculus) against SNP locations across chromosomes. The plot reveals a peak at the rs67827860 SNP located in an intron of *VCAN*. The lower grey line indicates a -$\log_{10}$(p-value) threshold of 7.5 and the upper one a threshold of 11. The other SNPs showing similarly high p-values (forming a dotted vertical line) are in LD with this SNP. b) Spatial mapping of the rs67827860 SNP against voxel-wise ICVF in white matter. Here, ICVF was averaged across the subjects having one copy of the reference allele and subtracted from the voxel-wise average across subjects having zero copies of the non-reference allele, with the difference thresholded at 0.05. c) The PheWAS plot plotting -$\log_{10}$(p-values) against all dMRI IDPs for the rs67827860 SNP, revealing notable associations across the spectrum of dMRI IDPs, including DTI-derived metrics. All figures adapted from Elliott and colleagues (Elliott et al., 2018) and added to the thesis with permission from the authors.

Importantly, the biological cause behind these associations remains elusive. MRI generally lacks specificity, and the identified MRI signals are not predicted to be directly sensitive to the extracellular matrix. One way we can begin to understand the cause of these changes is to conduct invasive comparisons of MRI phenotypes with histological investigations. This thesis proposes such a *reverse* translational approach

to population neuroimaging, starting with the *VCAN* and *BCAN* associations identified in UK Biobank imaging GWAS and designing hypothesis-driven MRI-histology studies in rodent models of these genes.

3.4. Basics of histology

Microscopy enables us to visualise components and characteristics of tissue in very high detail – up to a fraction of micron in resolution. Histology is the area of microscopy examining the micro-anatomy of tissue and is used in both research and clinical settings. Various stains are used to generate contrast, each providing different information about tissue microstructure. This section will discuss notions related to immunohistochemistry (IHC) of formalin-fixed, paraffin-embedded (FFPE) brain tissue specimens, the histological technique used in this thesis. The major advantages of FFPE-IHC are the preservation of structural morphology and its long-term storage capability.

The basic histology protocol for formalin-fixed, paraffin-embedded samples is highlighted in **Figure 3.6**. The histology protocol starts with the dissection of the tissue of interest. In the case of *ex-vivo* mouse brain tissue samples, the brains are extracted from the skull, ensuring minimal damage.

Following this, if only interested in a small area of the brain, e.g., one hemisphere, the brains can be sectioned further.

This is followed by processing, a stage where the tissue is taken from its fixation state to a state where it is infiltrated by another substance to facilitates its slicing, such as wax. To do that, the tissue section of interest is passed through alcohol solutions of increasing concentrations, in order to remove the water and unbound fixative from the tissue. This is followed by immersion in a clearing solution (e.g., citroclear), which serves as intermediary between the alcohol and wax, removing alcohol and making the tissue receptive to wax infiltration. Paraffin wax conserves the morphology of the

sample, provides optimal support during slicing and allows for long-term storage. Finally, the tissue is immersed in wax for a period of time, and then embedded in a block of wax. Processing and embedding can be manual, automatic or a mixture of both. The timing of the processing and embedding steps severely impacts the quality of the acquired tissue slices. Poorly processed tissue can result in holes, tears and folds or tissue breaks during water bath floating. Therefore, the processing and embedding stage should be optimised according to the characteristics of the tissue. Aspects such as the size of the tissue section, the duration of post-fixation storage of the samples or the tissue fixation protocol (perfusion vs immersion) need to be considered.

Following this, the tissue is very thinly sliced (5-10μm) using a microtome, and briefly floated in a water bath usually set at 10°C below the melting point of paraffin. The section is then manually picked up on a suitably sized slide and left to dry overnight. There are various solutions which can be used before or after slicing to obtain high quality slices, such as using a softening agent (e.g., Mollifex) before slicing, floating in ethanol after slicing and before placing the slice in the water bath or adjusting the slice thickness.

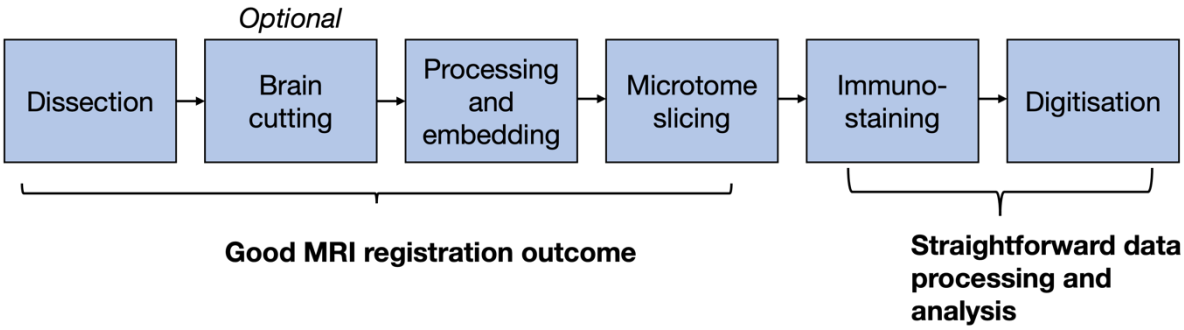

Figure 3.6. Steps for the formalin-fixed, paraffin-embedding processing protocol for rodent brain samples. Optimising the first four steps improves the quality of registration to MRI. The latter two steps need to be optimised to facilitate the extraction of quantitative metrics: reducing staining and digitisation artifacts significantly reduces the amount of time needed to process this data and extract reliable and replicable quantitative estimates.

The slices can be stored at room temperature for long periods of time prior to immunostaining. IHC is an application of immunostaining that labels proteins in tissue using specially-designed “primary” antibodies with an affinity for part of a particular protein of interest. “Secondary” antibodies bind to primary antibodies to generate signal amplification. The secondary antibody also carries a “chromogen” to create a coloured precipitate at the location of the protein of interest, enabling the visualisation of its location. The most common chromogen used in FFPE-IHC is 3,3α-diaminobenzidene tetrahydrochloride (DAB). A counterstain using a different chromogen is often used to visualise the location of the DAB with respect to individual cells. Haematoxylin is a common counterstain chromogen, colouring the nuclei of the cells purple. An example FPE-IHC myelin-stained sagittal mouse brain slice is shown in **Figure 3.7**.

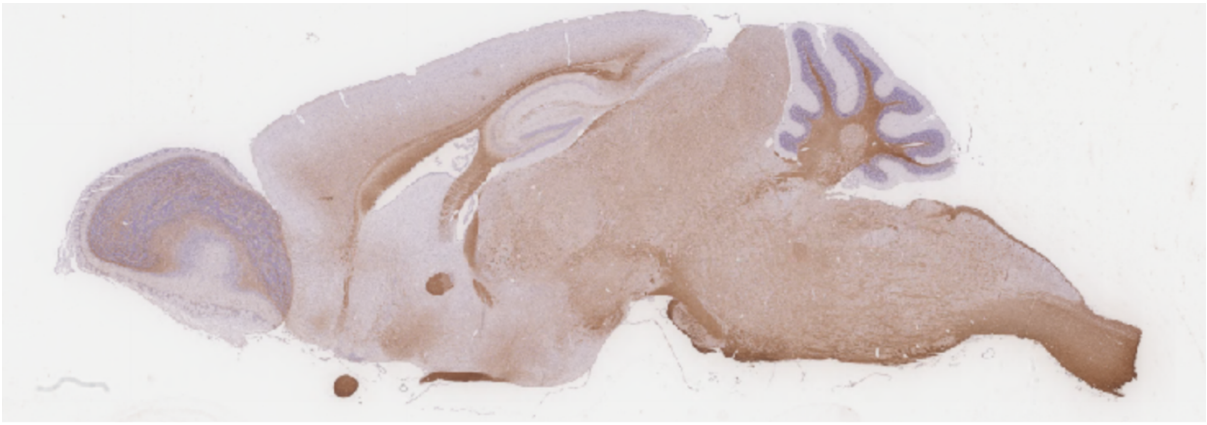

Figure 3.7. PLP stained sagittal mouse brain slice, counterstained with haematoxylin. Image courtesy of Aurea Martins-Bach *(unpublished data)*.

Common stains for brain histology include those targeting myelin proteins, e.g., myelin proteolipid protein (PLP) or myelin basic protein; neurofilaments, protein polymers found in neuronal cytoplasm and particularly abundant in axons and dendrites; neuronal nuclei; ferritin, an iron-storing protein; glial fibrillary acidic protein, an astrocyte marker or cluster of differentiation 68, a microglia and macrophages marker. These stains have been correlated tio MRI metrics in various studies (Chandran et al.,

2012; Chang et al., 2017; Hametner et al., 2018; Mollink et al., 2019; C. Wang et al., 2020b; Kor et al., 2022).

Following staining, the data is digitised, usually using a slide scanner which acquires data at sub-micron resolution, approximately three orders of magnitude higher than the typical rodent or human MRI dataset.

3.5. MRI-histology data fusion

3.5.1. MRI-histology registration

The ever-increasing accuracy of MRI-histology registration tools is bringing about massive improvements in linking these modalities. In theory, any coronal, axial or sagittal brain slice, as well as smaller tissue sections, can be registered to MRI. MRI-histology registration can be done either using serial sections (via their volumetric reconstruction) (Pichat *et al.*, 2018) or single sections, where the registration is done for each slice at a time. MRI-histology registration tools generally address important challenges such as tissue deformations and shrinkage or slice thickness mismatch between MRI and histology (Alyami, Kyme and Bourne, 2022). Importantly, it is necessary to conduct this registration automatically with minimal manual input.

FSL's Tensor Image Registration Library (TIRL) (Huszar *et al.*, 2023) supports histology-to-MRI (2D to 3D) and histology-to-histology (2D to 2D) registration of individual brain slices. TIRL is designed to allow for the registration of sparsely-sampled, single-section histology images to MRI. TIRL is a multi-stage registration pipeline requiring minimal manual input. The TIRL stages perform rigid, affine, in-plane and free-form deformations, using a modality-independent cost function.

3.5.2. Extraction of quantitative metrics

To correlate MRI and histology at the level of MRI voxels, we need to extract summary metrics from histological data that pool information over the same spatial extent as an MRI voxel. Software tools used in most neuropathological studies rely heavily on manual input (van der Weijden *et al.*, 2021). This can lead to inconsistencies across experiments and is time-consuming. In this thesis, we use an in-house automatic pipeline (Kor et al., 2022) to generate SAF maps for each stained slide. The SAF represents the number of DAB-stained pixels within a given area. Typically, square pixel patches are chosen (e.g., 15-25^2 in size). The main stages of the SAF calculation are separation of the DAB and haematoxylin stains using colour deconvolution; automatic segmentation of the DAB stain by identifying which pixels are above some intensity threshold (e.g., using Otsu's method); and calculation of the SAF as the fraction of pixels in a given area that exceed that threshold. To correlate this data to MRI, SAF across larger patches is averaged, matching the resolution of the MRI data.

There is strong evidence for good correlation between various microstructural MRI metrics and quantitative metrics extracted from histology (not necessarily SAF). A recent review estimates the presence of strong correlations between FA and radial diffusivity and myelin stains (Lazari and Lipp, 2021). Recent research has also shown positive correlations of QSM and R_2^* in human grey matter areas with the myelin PLP stain (C. Wang et al., 2020), as well as some positive trends linking NODDI metrics with histological metrics in humans (Grussu *et al.*, 2017).

Chapter 4: Multi-modal MRI in *ex-vivo* mouse brains: Development of acquisition and post-processing protocols comparable to UK Biobank

Context

This chapter describes the development of multi-modal MRI acquisition and post-processing protocols to phenotype *ex-vivo* mouse brains. The protocols include T_2-weighted structural MRI, diffusion-weighted MRI (dMRI) and multi-echo gradient echo (GRE), used for R_2^*- and quantitative susceptibility mapping (QSM). These were the first studies conducted in a new preclinical imaging facility, using a 7T Bruker scanner. We developed scanning protocols that aimed for alignment to the UK Biobank human MR neuroimaging study, enabling us to extract equivalent phenotypes from the rodent data. Although these new protocols and pipelines were developed for the specific project described in the thesis, they are now in use at our centre for investigations of stroke, neurodegeneration, and plasticity.

The code for our dMRI, R_2^*-mapping and QSM post-processing pipelines is openly available (https://git.fmrib.ox.ac.uk/preclinical-imaging), alongside guides on how to run them. **Appendix A** presents a step-by-step protocol on how to prepare the samples, run and evaluate the multi-modal MRI data acquisition and processing pipelines.

Published work

Personal contributions

Conference proceedings

Cristiana Tisca, Mohamed Tachrount, Frederik J Lange, Amy FD Howard, Chaoyue Wang, Lily R Qiu, Benjamin C Tendler, Jason P Lerch, Aurea B Martins-Bach, Karla L Miller. *White matter microstructure changes in a Bcan knockout mouse model.* ISMRM & ESMRMB Joint meeting (2022).

Cristiana Tisca, Mohamed Tachrount, Chaoyue Wang, Lily Qiu, Javier Barallobre-Barreiro, Marika Fava, Manuel Mayr, Jason Lerch, Aurea B Martins-Bach, Karla L Miller. *Vcan mutation leads to sex-specific changes in white matter microstructure in mice*. ISMRM & SMRT Virtual Conference (2021).

Cristiana Tisca, Amy F D Howard, Saad Jbabdi, Aurea B Martins-Bach, Karla L Miller. *Comparison of NODDI parameter estimates in staggered versus non-staggered 2-shell acquisitions.* ESMRMB, Virtual Conferences (2020).

Software

Cristiana Tisca, Mohamed Tachrount, Frederik J Lange, Amy FD Howard, Jason P Lerch, Aurea B Martins-Bach, Karla L Miller. *A diffusion-weighted MRI post-processing pipeline for ex-vivo rodent brains to extract DTI, DKI and NODDI metrics (1.0.0)*. Zenodo. https://doi.org/10.5281/zenodo.8129321.

Cristiana Tisca, Mohamed Tachrount, Frederik J Lange, Chaoyue Wang, Benjamin C Tendler, Jason P Lerch, Aurea B Martins-Bach, Karla L Miller. *R2*- and quantitative susceptibility mapping (QSM) post-processing pipelines for ex-vivo rodent brains (1.0.0)*. Zenodo. https://doi.org/10.5281/zenodo.8130910.

Collaborations

Articles

Florence Rawlings-Mortimer, Alberto Lazari, **Cristiana Tisca**, Mohamed Tachrount, Aurea B Martins-Bach, Karla L Miller, Jason P Lerch, Heidi Johansen-Berg. *7,8-dihydroxyflavone enhances long-term spatial memory and alters brain volume in wild-type mice.* Front. Sys. Neurosci (2023).

Merima Šabanović, Alberto Lazari, Marta Blanco-Pozo, **Cristiana Tisca**, Aurea Martins-Bach, Mohamed Tachrount, Jason P Lerch, Mark E Walton, David M Bannerman. *Lasting dynamic effects of the psychedelic 2,5-dimethoxyl-4-iodoamphetamine ((±)-DOI)) on cognitive flexibility.* Molecular Psychiatry (2024).

Conference proceedings

Jenna Hanmer, Martin Craig, Malcolm Prior, Joao N Correia, Matteo Bastiani, **Cristiana Tisca**, Aurea Martins-Bach, Mohamed Tachrount, Kurt G Schilling, Jason P Lerch, Paul S Morgan, Tracy D Farr, Stamatios N Sotiropoulos. *A pipeline for concurrent whole-brain 3D microscopy and MRI in the rat*. OHBM Annual Meeting (2023).

Aurea B Martins Bach, **Cristiana Tisca**, Mohamed Tachrount, Carmelo Milioto, Shoshana Spring, Mireia Carcolé, Remya R. Nair, Thomas J Cunningham, Elizabeth M C Fisher, Adrian M Isaacs, Brian J Nieman, Jason P Lerch, Karla L Miller. *Distinct patterns of microstructural changes in mouse models of ALS/FTD with mutations in Tardbp, FUS and C9Orf72.* ISMRM & ISMRT Annual Meeting (2023)

Kamila U Szulc-Lerch, Eugenio Graceffo, **Cristiana Tisca**, Aurea B Martins-Bach, Karla L Miller, Clemence Ligneul, Jason P Lerch, Yvonne Couch, Tracy Farr, Heidi

Johansen-Berg. *Mapping acute stroke in the mouse using structural and diffusion MRI*. Brain and Brain PET Annual Conference (2022).

Kamila Szulc-Lerch, Eugenio Graceffo, Lily Qiu, **Cristiana Tisca**, Aurea B Martins-Bach, Karla L Miller, Mohammed Tachrount, Claire Bratley, Clémence Ligneul, Antoine Cherix, Jason Lerch, Yvonne Couch, Tracy Farr, Heidi Johansen-Berg. *Multi-modal MRI Characterisation of a Mouse Model of Distal Middle Cerebral Artery Occlusion.* BRAIN Conference (2022).

Aurea B Martins-Bach, Mohamed Tachrount, Cristiana Tisca, Lily Qiu, Shoshana Spring, Jacob Ellegood, Brian J Nieman, John G Sled, Remya Raghavan-Nair, Elizabeth Fisher, Thomas Cunningham, Jason Lerch, Karla Miller. *Anatomical and microstructural brain alterations in the TDP-M323K mouse model of amyotrophic lateral sclerosis*. ISMRM & SMRT Virtual Conference (2021).

Individual contributions to this thesis chapter

My personal contributions to this chapter include conducting SNR tests in phantoms; designing the sample preparation for *ex-vivo* MRI and quality control protocols; development of the multi-modal MRI acquisition protocols; addressing image artifacts; conducting experiments investigating angular resolution for the dMRI acquisition; developing dMRI, R_2^*- and QSM post-processing pipelines.

The protocol optimisation and pipeline development work for all sequences was collaborative, with additional input from Aurea Martins-Bach and Mohamed Tachrount. The contribution of other collaborators is listed in the **Acknowledgements** section.

4.1. Introduction

MRI enables the non-invasive phenotyping of *ex-vivo* rodent brains, including novel, genetically-modified mouse models. Genetically-modified mouse models are ideal for studying the effect of genetic variation on brain structure and function. Using MRI protocols harmonised to human imaging, we can observe equivalent phenotypes in mouse brains. Furthermore, preclinical MRI scanners benefit from higher field strengths, stronger gradients and cryogenically-cooled coils which significantly boost SNR. In preclinical studies, we can characterise brains at higher resolution and acquire more extensive datasets.

This chapter presents our investigations and optimisations for implementing multi-modal MRI protocols for *ex-vivo* mouse brains that are comparable to the UK Biobank neuroimaging protocols. We acquire high-resolution structural and microstructural MRI data in a 14h overnight timeslot. This includes a 30min high-resolution structural MRI protocol using a TurboRARE sequence. The obtained data is used for volumetric analysis (Chakravarty *et al.*, 2013; Friedel *et al.*, 2014) and subsequent registration to histology. The 13-hour dMRI acquisition protocol used an edited 3D segmented EPI sequence. Its optimisation focused on the number of segments used, optimal b-values, number of directions and their configuration (staggered vs non-staggered). We developed a robust pipeline that deals with artifacts (Gibbs-ringing, susceptibility distortion correction) and fits the DTI and NODDI models. Finally, a 3D multi-echo GRE sequence was used for R_2^*-mapping and QSM. The pipelines for producing these parametric maps were based on recently-published work (C. Wang et al., 2020). Each modality required identification and eradication of image artifacts. The pipelines are openly shared online (https://git.fmrib.ox.ac.uk/preclinical-imaging).

Finally, we built a custom holder, setup and preparation protocol for mouse brains. This is accompanied by a quick, 30min protocol to assess sample integrity before the 14h timeslot.

These protocols and pipelines represent the first work conducted on a new 7T Bruker BioSpec 70/20 USR scanner. The sequence optimisation experiments and pipeline development took about 3-4 months each.

Following the optimisation of the protocols, we acquired and processed this multi-modal MRI data in our two cohorts, *Vcan* and *Bcan*. The cohorts and study results will be presented in **Chapter 6**.

4.2. Data acquisition

4.2.1. Hardware considerations

Our initial work sought to identify the optimal scanning setup for *ex-vivo* mouse brains using strong gradient coils of up to 660mT/m. We conducted SNR tests of several Bruker receive coils (**Table 4.1**) using a 3D EPI sequence (12 segments, TE=27ms, BW=476kHz, b=0s/mm^2, matrix size: 220x250x220, resolution: 0.1mm^3 isotropic). These tests used a Gadolinium-based phantom (Gadovist, 2mmol/L, T_1=88ms, T_2=75ms). SNR was calculated as the average signal in the region of interest divided by the standard deviation of the background noise in the region of interest. Tests were done using the same field-of-view (FoV), covering approximately the same phantom region. Due to the significantly higher SNR of the cryoprobe, we opted to use it and scan one sample at a time. The cryoprobe is a surface coil with 4 channels.

Coil	Coil type	Approximate cross-sectional SNR	Number of samples which can be scanned simultaneously
Cryoprobe	Surface	131.87	1
23mm	Volume	86.97	1
35mm	Volume	49.04	2-3
40mm	Volume	40.90	2-3

Table 4.1. Comparison of SNR profiles between various coils tested.

4.2.2. Protocol development

In the following sections we will describe how we optimised protocols that provide imaging phenotypes aligned to the UK Biobank, using a 14h overnight timeslot. The acquisition details are presented in **Table 4.2**. We did not acquire equivalent T_1-weighted data as we needed only one structural imaging dataset for volumetric analysis. The T_2-weighted data showed comparatively improved grey-white matter contrast in our Gd-doped mouse brain samples. All our acquisitions use signal excitation and reception over a 3D slab to boost SNR and enable isotropic resolution (Schilling et al., 2022).

The protocol optimisation considered resolution, SNR, grey-white matter contrast and biophysical model fitting. Because our data analysis is focused on white matter (as will be described in **Chapter 6**), we aim for a voxel size that minimises partial volume effects while still maintaining good SNR. The final protocol uses an isotropic 100μm isotropic voxel size for the dMRI and multi-echo GRE acquisitions and 60μm isotropic voxel size for the structural data.

MRI modality	WIN *ex vivo* mouse			UK Biobank		
	Duration	**Voxel, Matrix**	**Key Parameters**	**Duration**	**Voxel, Matrix**	**Key Parameters**
T1-weighted	N/A	N/A	N/A	5min	1mm isotropic 208x256x256	3D MPRAGE, sagittal, R=2, TI/TR=880/2000ms
T2-weighted	33min	60µm isotropic 400x160x200	3D TurboRARE, TE=12ms, echo spacing=12ms, 6 echoes, TR=350ms, BW=60kHz. Pre-scan adjustments: Scan shimming repeated before each scan.	6min	1.05x1x1mm 192x256x256	FLAIR, 3D SPACE, sagittal, R=2, PF 7/8, fat sat, TI/TR=1800/5000ms, elliptical
dMRI	13h	100µm isotropic 240x96x120	3D segmented EPI, 12 segments, TE=30ms, TR=500ms, δ/Δ =6.7/13.5ms, b=0 s/mm^2 (volumes: 4+1 phase-encoding reversed), b=2,500 s/mm^2 (volumes: 30), b=10,000 s/mm^2 (volumes: 30), BW=250kHz, directions covering the whole shell and staggered between shells.	7min	2mm isotropic 104x104x72	MB=3, R=1, TE/TR=92/3600ms, PF 6/8, fat sat, b=0 s/mm^2 (volumes: 5 + 3 phase-encoding reversed), b=1,000 s/mm^2 (volumes: 50), b=2,000 s/mm^2 (volumes: 50), staggered directions between shells
T2*-weighted	26min	100µm isotropic 240x96x120	3D Multi-echo GRE, FA=15°, TE1=3ms, echo spacing=3ms, 20 echoes, TR=68ms, NR=2, BW=250kHz, monopolar readout, prospective navigator correction. Pre-scan adjustments: Scan shimming repeated before each scan.	2min 30s	0.8x0.8x3.0mm 256x288x48	3D GRE, axial, R=2, PF 7/8 TE1/TE2/TR=9.4/20/27ms, NR=2

Table 4.2. Acquisition parameters for the *ex-vivo* mouse brain protocol and comparison with the UK Biobank protocol. TurboRARE: Turbo Rapid Acquisition with Relaxation Enhancement; GRE: gradient-echo; EPI: Echo-Planar Imaging; MPRAGE: Magnetisation-Prepared Rapid Acquisition Gradient Echo; FLAIR: Fluid-Attenuated Inversion Recovery; TE: echo time; TR: repetition time; BW: bandwidth; FA: flip angle; TE1: first echo time; NR: number of repetitions; δ: duration of diffusion-weighted gradient; Δ: time interval between the start of the first gradient and the start of the second gradient; ATS: Animal Transport System; R: acceleration factor; PF: partial Fourier; TI: inversion time; MB: multiband.

4.2.2.1. Structural T_2-weighted MRI

The T_2-weighted acquisition uses a modified 3D TurboRARE protocol available from Bruker (**Table 4.2**). A visual comparison was conducted between datasets acquired at different resolutions to determine the sequence details offering the optimal grey-white matter contrast and SNR, both of which affect the outcome of image registration. The other sequence parameters we used were optimised by our collaborators at the University of Toronto for their T_2-weighted sequence. Increasing the acceleration factor (number of phase-encoded echoes) can reduce the acquisition time at the cost of increasing the effective TE (the time between excitation and traversal of the k-space centre) and additional T_2-blurring.

This data was used for registration to a cohort-specific common space (Friedel *et al.*, 2014), to the DSURQE mouse brain atlas (Chakravarty *et al.*, 2013) and to histological data (Huszar *et al.*, 2023).

I observed an artifact when using the initial protocol which resulted in blurred, unusable images (**Figure 4.1**). We ran test acquisitions without phase-encoding and detected a phase offset across the echoes, which was reduced following adjustments to the default spoiling parameters. Alterations include: Auto Repetition Spoiler between echoes: ON → OFF, spoiling: 4 cycles → 1 cycle, BW=100 → 60kHz. Sequence simulations done in MATLAB using the Shepp-Logan phantom and the phase offsets derived experimentally revealed similar artifact patterns. Both our subsequently-acquired data and the Shepp-Logan phantom experiments showed significant improvements following the implemented modification (**Figure 4.1**).

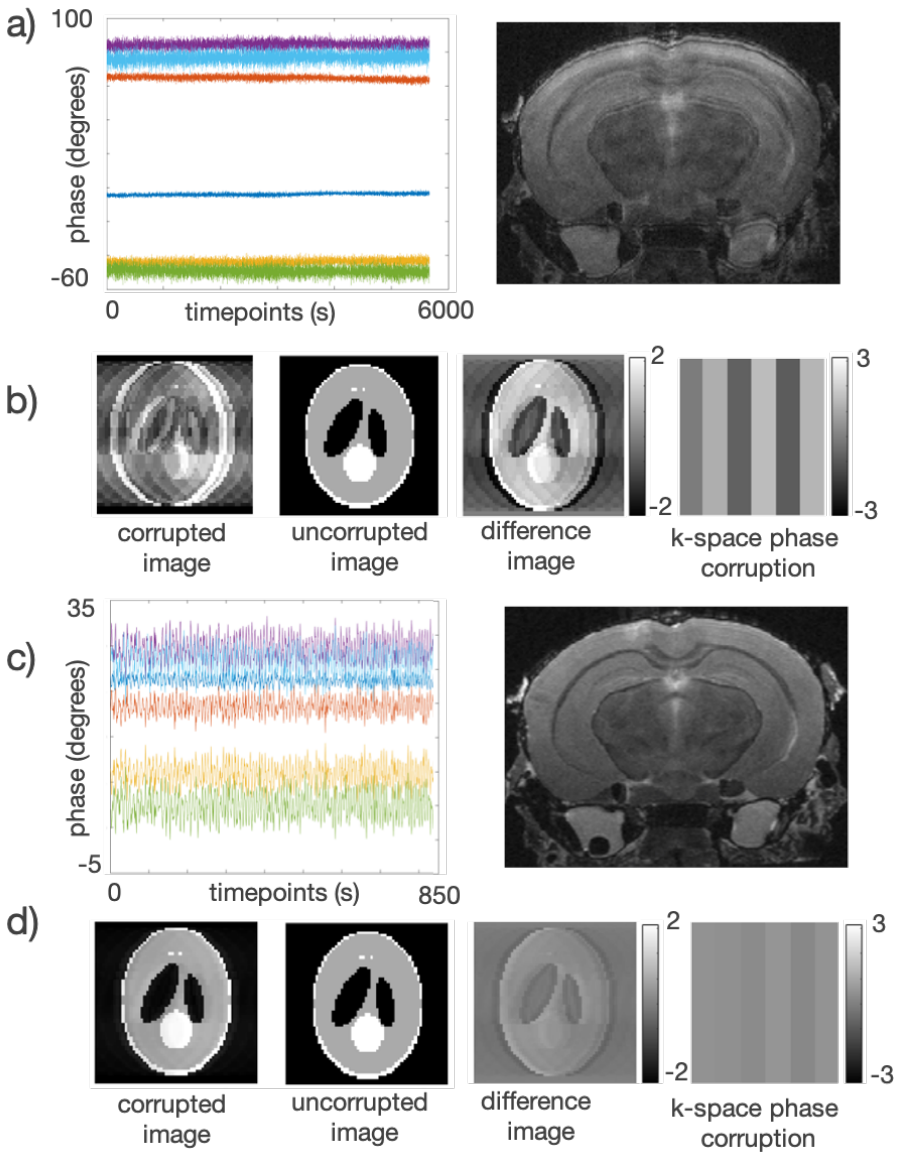

Figure 4.1. a) Left: Phase offset across the 6 echoes (each in a different colour). Acquisition without phase encoding. Right: T_2-weighted image acquired with the default spoiling parameters, showing artifacts throughout the image. b) Simulations of the artifact using phase values derived experimentally (shown in a), displaying a similar artifact. c) Left: Phase offset significantly reduced across the 6 echoes (each in a different colour) following the modifications to the default spoiling parameters. Acquisition without phase encoding. Right: T_2-weighted image acquired with the modified spoiling parameters, without any major artifacts present. d) Simulations of the resulting image using phase offset values derived experimentally following the modification to the default spoiling parameters. Note the difference in scale between plots a) and c).

4.2.2.2. Diffusion-weighted MRI

The dMRI sequence uses a 12-segment 3D segmented EPI sequence from Bruker (**Table 4.2**). Each 3D volume is acquired in equally spaced, interleaved segments of k-space. Segmented EPI presents reduced distortions compared to single-shot EPI.

We investigated the optimal combination of b-values. It is common to fit DTI to b-values of up to 1,000s/mm^2 *in-vivo*. Due to reductions in T_1, T_2, T_2^* and diffusivity in *ex-vivo* tissue, higher b-values are required (Roebroeck, Miller and Aggarwal, 2019) to achieve equivalent diffusion contrast. Further, to extract similar NODDI phenotypes to UK Biobank ones, we acquired data at two shells. To determine the optimal second shell value, we acquired datasets at b=2,500; 5,000; 7,000 and 10,000 with 30 uniformly-dispersed directions across each shell and determined which combination of b-values yielded minimum variance on NODDI parametric maps. We identified b=2,500 and b=10,000mm/s^2 as the optimal choice, where the least amount of noise was observed on the estimated parametric maps, especially on the intra cellular volume fraction (ICVF) maps. This is in line with the results in the original NODDI paper (Zhang *et al.*, 2012), where the lowest error in volume fractions was achieved for the protocol with the widest range of b-values.

Gradient directions were uniformly distributed across the whole shell using FSL GPS. We interleaved b=0s/mm^2 every 15 volumes to prevent overheating of the gradients and enable signal drift monitoring and correction across volumes. We also acquired a volume using a reversed phase encoding direction ("Blip Down"), used as part of the input to TOPUP (Andersson, Skare and Ashburner, 2003) and EDDY (Andersson and Sotiropoulos, 2016). In total, we acquire 30 diffusion-weighted volumes at each shell, which is enough to obtain a good signal fit to our models of interest and minimise the effects of noise.

I investigated whether using staggered (i.e., having different orientations) or non-staggered (i.e., having the same orientation) gradient directions across the two shells had an impact on the NODDI-derived parametric maps. I conducted a short study

(Tisca *et al.*, 2020) using an extensive *ex-vivo* macaque dataset (Howard et al., 2023) (**Figure 4.2**) containing 250 directions. We found no statistically significant differences between staggered and non-staggered acquisitions in NODDI's OD, ICVF or ISOVF maps across a white matter mask. However, ISOVF estimates tended to improve in the staggered approach using at least 30 directions per shell. This motivated our choice for staggering directions across shells.

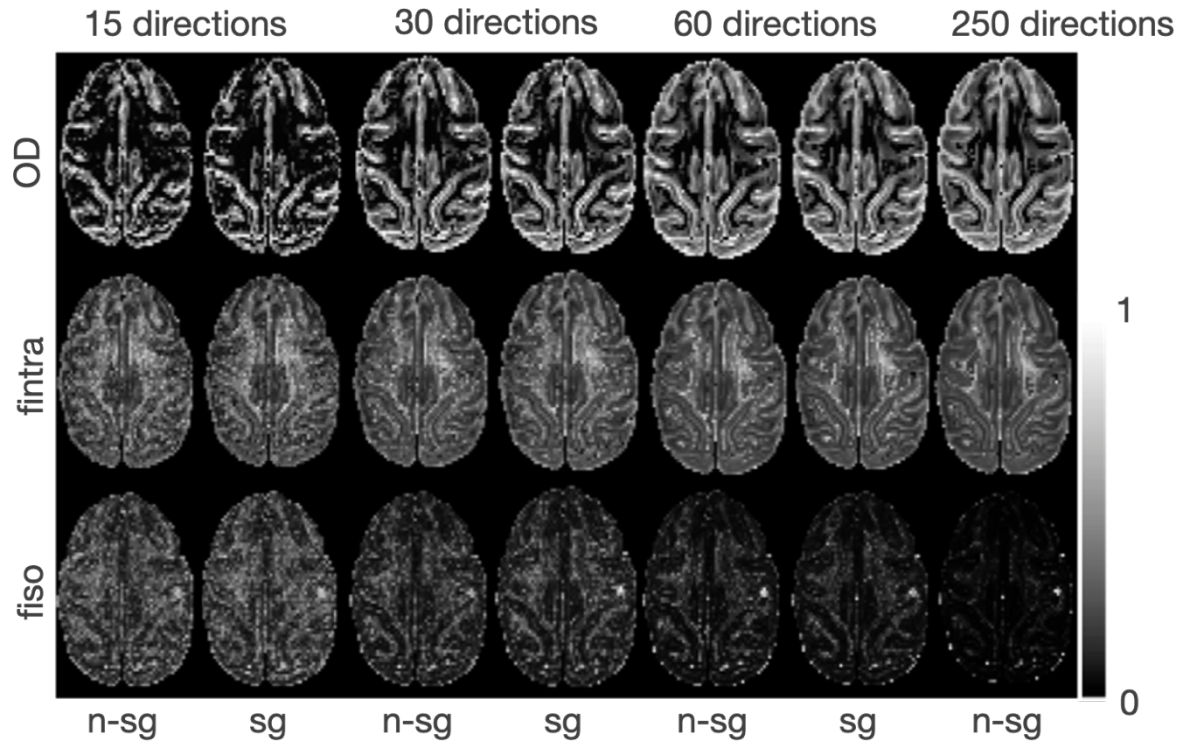

Figure 4.2. Example NODDI parametric maps obtained from the BigMac dataset (Howard et al., 2023), obtained using staggered (sg) and non-staggered (n-sg) directions across the two shells (b=7,000 and 10,000 mm/s^2) and either 15, 30 or 60 gradient directions. Note that the 250 gradient directions dataset only contained non-staggered directions across the two shells.

Following the optimisation of the acquisition parameters, I observed a drop in signal intensity across all volumes (**Figure 4.3**, a, red arrow) on one side of each imaging volume. This was attributed to B_0 inhomogeneities distorting the profile of the excitation RF pulse and affecting the signal intensity of thinner slices within the

excitation slab (in our case, on the left side of the image). To avoid these effects, we increased the thickness of the excited slab while keeping the encoded FoV constant. We experimentally determined that a 1.5 excited slab/FoV ratio eliminated this problem (**Figure 4.3**, b). The ability to adjust this ratio was incorporated as a select-option element into the sequence.

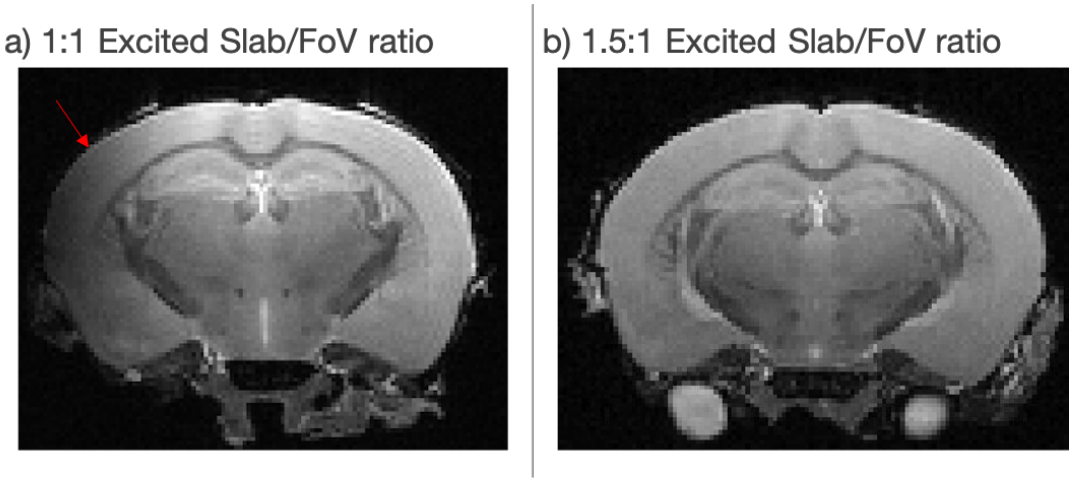

Figure 4.3. a) Signal drop-out observed on one side of the image when using a 1:1 excited slab/FoV ratio. b) No signal dropout observed when increasing the excited slab/FoV ratio to 1.5:1.

4.2.2.3. Multi-echo gradient-echo

The multi-echo GRE data is acquired at 100μm isotropic resolution using a standard sequence, comprising a slice-selective excitation followed by multiple gradient echoes resulting from switching of the readout gradient, all echoes being identically phase-encoded.

Our data is acquired using monopolar readout to reduce visible misalignment between data acquired at odd and even echoes (Cohen-Adad, 2014). During optimisation, we determined the minimum number of echoes that would sample the signal profile uniformly until the noise floor is reached, as well as achieve accurate phase contrast. Using 20 echoes satisfied this condition. Reducing the TE to 2ms lead to a critical duty cycle of 94.2%. To avoid that, we used a TE of 3ms.

We identified severe blurring across each volume and misalignment between volumes acquired at different echo times (**Figure 4.4**). The artifact occurred due to frequency drift intrinsic to our MR system design. It resulted from slow heating of gradient set which subsequently heated up the passive shim. This was solved using prospective navigator correction, where we apply a low flip angle, non-selective excitation pulse before each excitation. The frequency drift is estimated from the resulting spectrum and corrected during the scan.

We also increased the number of dummy scans acquired before the scan to the maximum value, allowing the temperature to stabilise. This sequence has a maximum duty cycle of 50.3%, compared to 13.4% achieved at most during the dMRI acquisition. The high duty cycle is due to the short TR and high number of echoes acquired per TR. Due to steep temperature rise that occurs at the beginning of this sequence, it is acquired last. The average temperature across all gradient coils during stand-by is approximately 16°C. During the multi-echo GRE acquisition, this value rises to around 22°C, compared to an average of 18°C recorded during the dMRI acquisition.

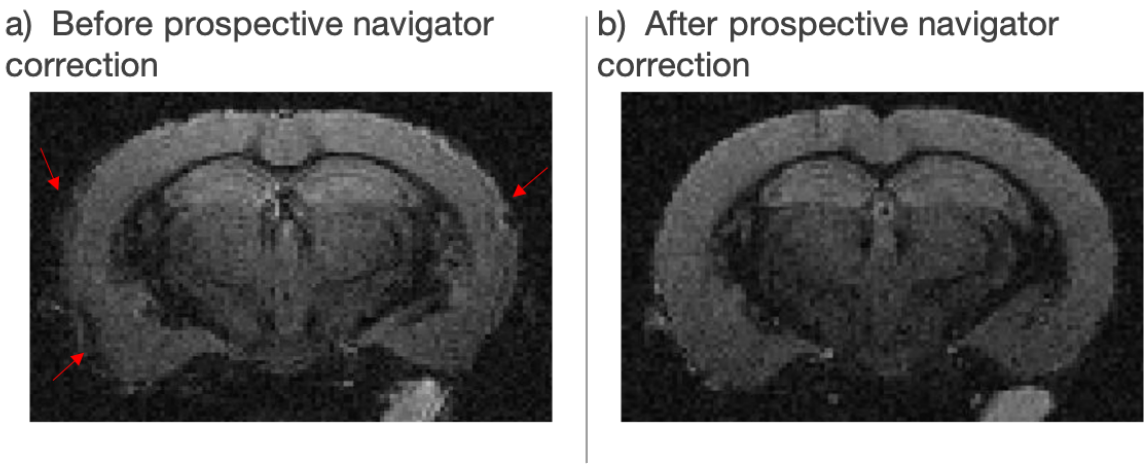

Figure 4.4. a) Artifact present in the Multi-echo GRE data (echo 11). Blurring is present throughout the image, with signal non-uniformities around the edge of the brain (red arrows). b) Data after implementation of the prospective navigator correction.

4.3. Sample preparation and scanning

4.3.1. Perfusion fixation

Our mouse brain samples are obtained following an established perfusion protocol (Cahill *et al.*, 2012) (**Figure 4.5**). Following perfusion-fixation and 24h in PFA, the samples are stored for long periods of time in a solution of phosphate-buffered saline (PBS), Gadolinium-based contrast agent (Gadovist, 2mM) and sodium azide 0.01% at 4°C. Sodium azide inhibits fungal and bacterial growth. The long-term storage solution does not have fixative in it. Keeping the samples in fixative long-term will increase the number of cross-links between molecules and affect the relaxation properties of the samples (de Guzman *et al.*, 2016, Tendler *et al.*, 2021). 24h in fixative ensures fixation without a major impact on relaxation properties.

To ensure that the mouse brains are of adequate quality, it is important to use a constant 1ml/min flow rate and liquid PFA during perfusion fixation. The brains are kept in the skull following perfusion fixation to avoid any damage and preserve their morphology.

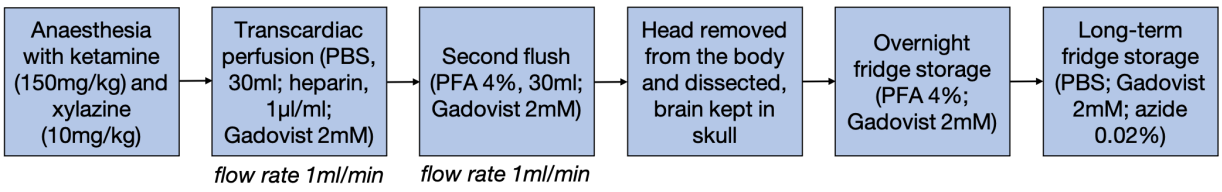

Figure 4.5. Sample perfusion and storage protocol. PBS: phosphate-buffered saline; Gadovist: Gadolinium-based contrast agent, PFA: paraformaldehyde.

4.3.2. Sample loading

Prior to MRI scanning, each sample is removed from the storage solution and positioned on a 3D-printed custom holder (**Figure 4.6**). The holder is placed in a falcon tube filled with Fluorinert (perfluorocarbon), a proton-free, MR-invisible substance

which does not damage sample quality (Iglesias *et al.*, 2018). The sample is then inserted in a vacuum pump for 30-60min (pressure: 10-20mbar) to remove any air bubbles stuck between the brain and skull. Before scanning, a bubble trap is placed in the falcon tube to block remaining air bubbles from reaching the surface of the skull. The sample is kept at room temperature for at least 1h prior to scanning to avoid temperature fluctuations during the scan. The falcon tube is loaded into the cryoprobe cradle in horizontal position and fixed in place using 3D printed custom parts to avoid any movement.

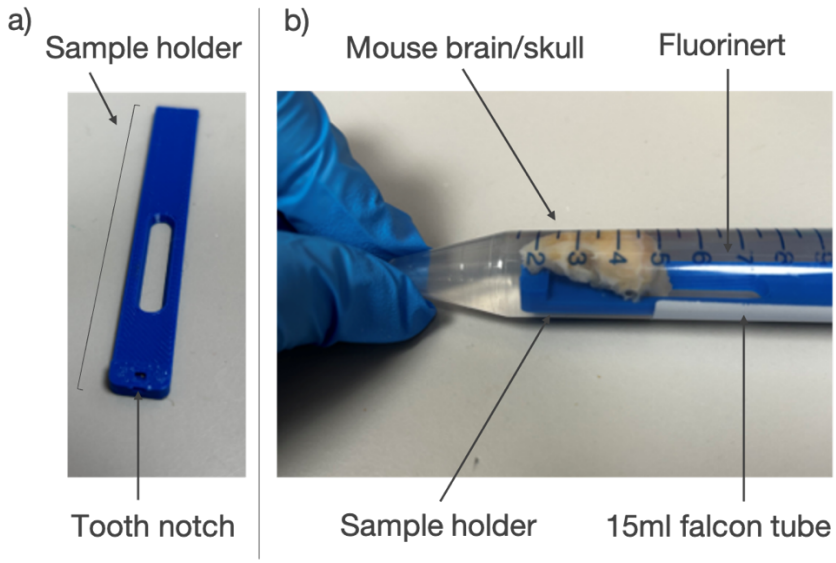

Figure 4.6. Photographs of the custom 3D-printed sample holder used for *ex-vivo* mouse brain samples. a) 3D-printed sample holder; b) Mouse skull placed on the sample holder using the tooth notch and fixed inside the falcon tube filled with Fluorinert.

4.3.3. Sample tests

If poorly perfused – not at the correct flow rate or using powder perfusates – or stored inadequately, mouse brain samples may deteriorate. Researchers have estimated that about 30% of brains prepared by perfusion have MR-visible artifacts (Cahill *et al.*, 2012). These include reduced grey-white matter contrast on T_2-weighted images, higher or inhomogeneous T_2 values across the brain, grey-white contrast inversion on

T_2 maps and a white, hyperintense rim at the edge of the brain, likely where the perfusates have accumulated. If the samples are accidentally frozen, small hyperintensities can be observed throughout the brain.

We developed a quick, 30min-long test protocol to assess sample quality (**Table 4.3**) prior to the overnight protocol. This protocol assesses gross neuroanatomy across the sample, enabling the detection of perfusion fixation-related artifacts. Our compromised samples presented the same artifacts as those described in literature. We also acquire a multi-echo GRE short scan to evaluate whether air bubbles are present in the skull. Data from a good-quality sample and a faulty one is shown in **Figure 4.7**.

MRI modality	Duration	Voxel size, Matrix	Acquisition details
T2-weighted structural	17min	80µm isotropic, 300x120x150	3D TurboRARE, TE=36ms, echo spacing=12ms, 6 echoes, TR=350ms, NR=1, BW=60.6kHz
T2-mapping - sagittal	6min each	133x133x500µm 180x75	2D Multi-slice multi-echo, TE=5ms, echo spacing=5ms, 25 echoes, TR=2500ms, NA=2, BW=100kHz, 18 slices
T2 mapping - axial			
T2*-weighted	2min	200µm isotropic, 120x48x55	3D Multi-echo GRE, TE=3ms, echo spacing=3ms, 10 echoes, TR=50ms, FA=15°, BW=250kHz

Table 4.3. Protocol details of the test scan conducted to assess sample integrity. The T_2 maps were computed and displayed using ParaVision 360 software. TurboRARE: Turbo Rapid Acquisition with Relaxation Enhancement; GRE: gradient echo; TE: echo time; TR: repetition time; NR: number of repetitions; NA: number of acquisitions; BW: bandwidth.

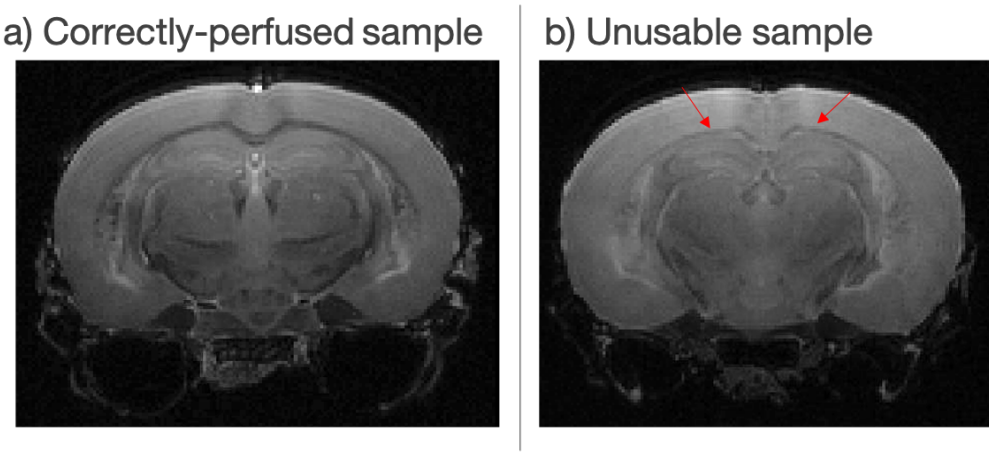

Figure 4.7. Example T_2-weighted structural images from a well-perfused sample (a) and a faulty one (b). Red arrows indicate areas in the corpus callosum where the grey-white matter contrast is reduced.

4.3.4. Data acquisition

Following sample tests, the multi-modal overnight MRI protocol is acquired at constant temperature (22°C). The samples are positioned in the same location in the scanner for each acquisition.

4.4. Data processing

4.4.1. T_2-weighted MRI

The T_2-weighted data from each individual cohort is registered to a cohort-specific common space (Friedel *et al.*, 2014) and segmented based on the DSURQE mouse brain atlas (Chakravarty *et al.*, 2013). This data is used to assess whether the genetically-modified mouse models show changes in brain volume. An example processed T_2-weighted image (brain-extracted and Gibbs-unringed) from a wild-type mouse brain is shown in **Figure 4.8**.

This data is used to extract an imaging-derived phenotype (IDP) equivalent to the grey-white matter contrast extracted from the UK Biobank datasets. To do this, we obtain a mask of the isocortex by non-linearly registering our data to the Allen Mouse Brain Common Coordinate Framework version 3 3D reference atlas (Q. Wang *et al.*, 2020). This is done using the output of the standard space cohort-wise registration (Yee *et al.*, 2018). For each grey matter voxel in our eroded isocortex mask, we identify the white matter voxel which is at the nearest Euclidean distance to it and take their ratio. The isocortex mask was eroded to avoid partial volume effects.

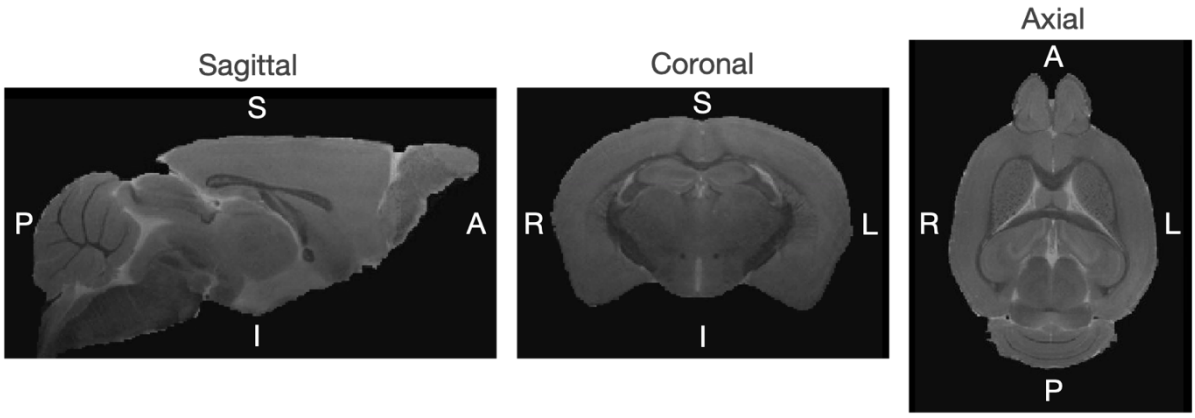

Figure 4.8. Example T_2-weighted MRI data from a wild-type mouse brain from the *Bcan* cohort. The data was corrected for Gibbs-ringing and the skull was stripped using the atlas-based segmentation.

4.4.2. Diffusion-weighted MRI

To be able to extract reliable quantitative metrics from our dMRI data, we employ a sequence of standard processing steps. Our data presents minimal artifacts – e.g., small susceptibility-induced distortions and sub-voxel shifts due to signal drift across volumes –thus requiring minimal pre-processing. Our processing pipeline accounts for particularities of Bruker data, is generalisable and easy to use. The pipeline steps are highlighted in **Figure 4.9**.

Having observed Gibbs ringing on all images near tissue boundaries, we apply the correction to all 64 volumes. We use a 3D extension to the 2D Gibbs ringing correction algorithm based on local sub-voxel shifts (Kellner *et al.*, 2016; Bautista *et al.*, 2021), available as part of the MRtrix3 software (https://mrtrix.readthedocs.io). This removes some residual ringing present in the readout direction as well.

Differences in magnetic susceptibility (χ) across air-tissue boundaries (e.g., where small microbubbles are trapped between the skull and the brain) lead to signal distortion observed across volumes. We use FSL's TOPUP (Andersson, Skare and Ashburner, 2003; Smith et al., 2004) to estimate the χ-induced field. The first b=0 volume, together with a volume acquired using a reverse phase encoding direction ("Blip Down") are used as TOPUP inputs. To be able to run TOPUP on high-resolution rodent data, we wrote a custom TOPUP configuration script where we modified the parameter values (resolution of warps, full-width half-maximum of Gaussian smoothing) used during the optimisation stage. We feed TOPUP's output into the EDDY tool (Andersson and Sotiropoulos, 2016), which corrects the distortions and any misalignment between volumes (in our case, due to signal drift instead of movement).

DTI can be fit to the lower b-value shell data (b=2,500mm/s^2) using FSL's DTIFIT, which is part of FMRIB's Diffusion Toolbox (FDT). Example parametric maps (FA, MD) are featured in **Figure 4.9**.

Finally, NODDI (Zhang *et al.*, 2012) is fit to our data using cuDIMOT (Hernandez-Fernandez *et al.*, 2019). One of the limitations of NODDI is the use of fixed axial (d_{ax}) and isotropic diffusivities throughout the brain. Literature values are often used in *ex-vivo* studies without accounting for how diffusivities change during perfusion fixation. Instead, we employed a data-driven approach to estimate d_{ax}. We used the spherical mean technique (SMT) (Kaden *et al.*, 2016) and a dataset acquired using our diffusion protocol. The SMT-estimated voxel-wise d_{ax} was fed into an adapted version of NODDI. The estimated signal from NODDI was used as an updated input to SMT. 10 iterations of this process were needed for the estimates to plateau. The mode of the histogram distribution of d_{ax} values across the corpus callosum was chosen as our

data-driven d_{ax} estimate for all dMRI datasets acquired using the same protocol: $d_{ax}=4.7x10^{-4}mm^2/s$. The isotropic diffusivity value, d_{iso} was set to 10^{-3} mm^2/s. A dot compartment was included for the NODDI fitting as well. The dot compartment models small spherical components of tissue, such as cell bodies, where water diffusion is isotropically restricted. In *ex-vivo* scenarios, the dot compartment models water trapped in tissue due to fixation.

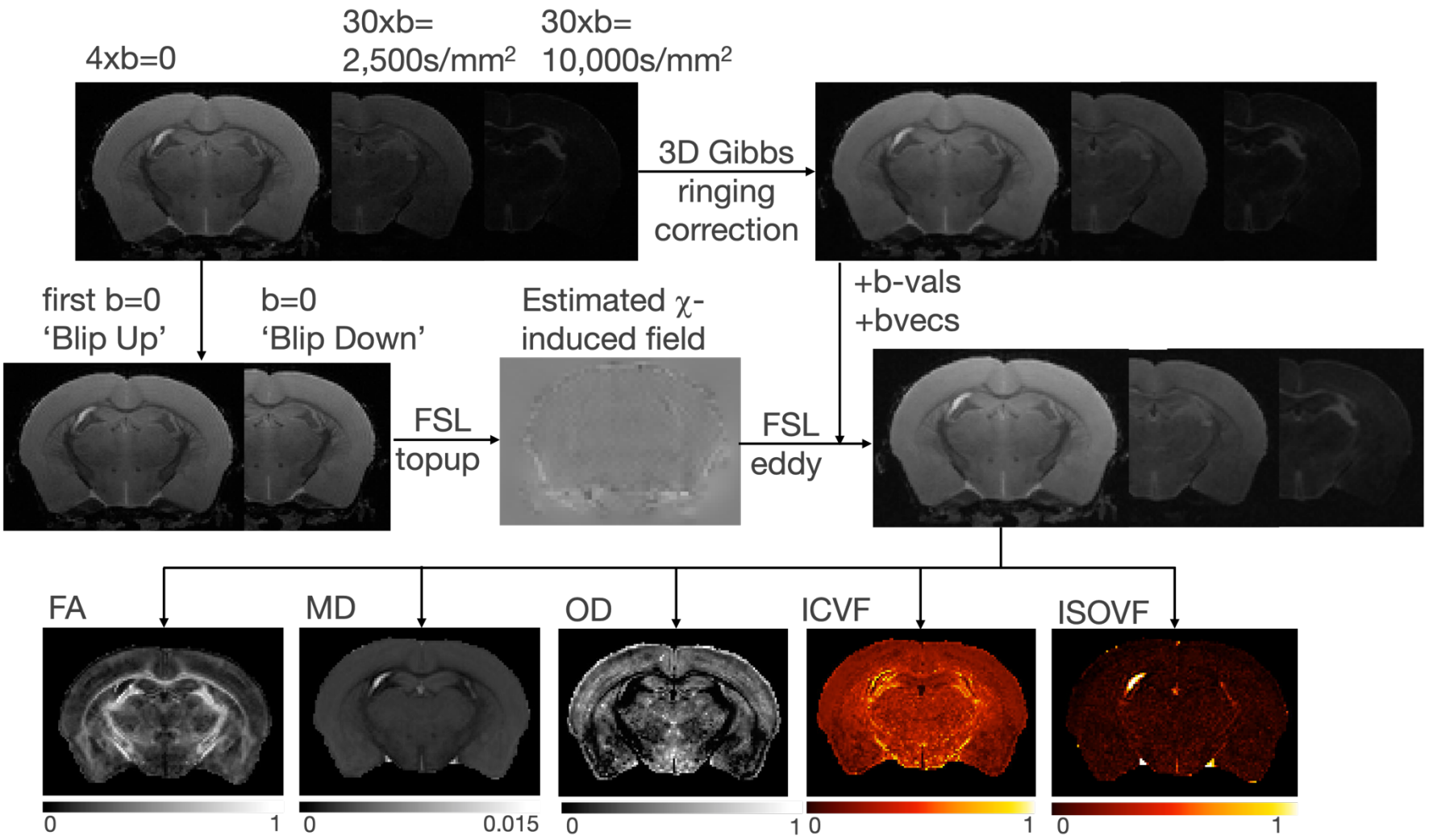

Figure 4.9. Diagram of the dMRI processing pipeline. b-vecs: b-vectors; b-vals: b-values; FA: fractional anisotropy; MD: mean diffusivity; OD: orientation dispersion; ICVF: intra-cellular volume fraction; ISOVF: isotropic volume fraction.

4.4.3. R_2^* mapping and quantitative susceptibility mapping

We process multi-echo GRE data to produce QSM and R_2^* maps, by adapting our group's existing pipeline for human post-mortem data (C. Wang et al., 2020). The pipelines combine custom code with several toolboxes to perform pre-processing (e.g., Gibbs ringing and cross-echo alignment) and quantification and is presented in **Figure 4.11**. The pipelines are designed to account for particularities of Bruker data and be user friendly, running in two stages (pre-processing and quantification), each using a single command.

4.4.3.1. Artifact Correction

We use FLIRT to perform 6-DOF registration of the magnitude images across TEs. Images are smoothed prior to registration, and the volume at each TE is registered to the previous one. The first TE volume of the second repetition is registered to the first TE volume of the first repetition. This strategy outperforms direct registration to the first TE (as neighbouring volumes have more similar contrast), leading to improved overall registration results. The final alignment is performed on the non-smoothed real and imaginary components for each cryoprobe channel (n=4) using the concatenated transform matrices.

A small degree of Gibbs ringing is present in our magnitude data and is particularly visible on the low-TE volumes. We use 3D Gibbs ringing correction (Bautista *et al.*, 2021) for all volumes to minimise this artifact. This correction is done following averaging of the volumes from both repetitions. The Gibbs ringing-corrected data for each individual repetition is also used in the voxel-wise fitting of the complex data, reconstructed separately from Bruker 2dseq files using a custom MATLAB script.

4.4.3.2. $\mathbf{R_2^*}$ processing pipeline

The averaged, Gibbs ringing-corrected magnitude data was used for voxel-wise R_2^* fitting across the 20 echoes, by minimising:

$$argmin_{M_0,R_2^*,noise} \left\| M(TE) - \left[(M_0 e^{-TE*R_2^*})^2 + \varepsilon^2 \right]^{0.5} \right\|_2^2 \tag{4.1}$$

Here, $M(TE)$ is the magnitude data at each voxel, M_0 is the extrapolated signal at TE=0ms and ε is the estimated noise component. The fitting used MATLAB's *lsqnonlin* function. R_2^* was initialised from the first 3 echoes, by estimating it via linear regression from the logarithm of the magnitude data. M_0 was estimated from the obtained R_2^* and the mean magnitude and TE from the 3 echoes. Here, the noise parameter is incorporated and estimated within the fitting, assuming a Rician distribution (Gudbjartsson and Patz, 1995).

4.4.3.3. Quantitative susceptibility mapping processing pipeline

Following co-registration, the real and imaginary components of the complex data for all coil channels (n=4) are combined (**Figure 4.10**). The phase components of each channel present phase offsets depending on the coil sensitivity profile (Haacke *et al.*, 2015). To avoid the introduction of phase cancellations or unwanted artifacts, such as signal cancellation or open-ended fringe lines, appropriate phase offset removal and coil combination needs to be conducted. We use an algorithm which subtracts the phase at the first TE from all subsequent echoes, separately for each repetition (Robinson *et al.*, 2017). This technique assumes that, at very short TEs, the measured phase approximates the phase offset. This subtraction aligns the phase across the coils. The final phase image is calculated across channels as the phase of the complex sum. The magnitude image is reconstructed using the sum of squares of individual channel magnitude images.

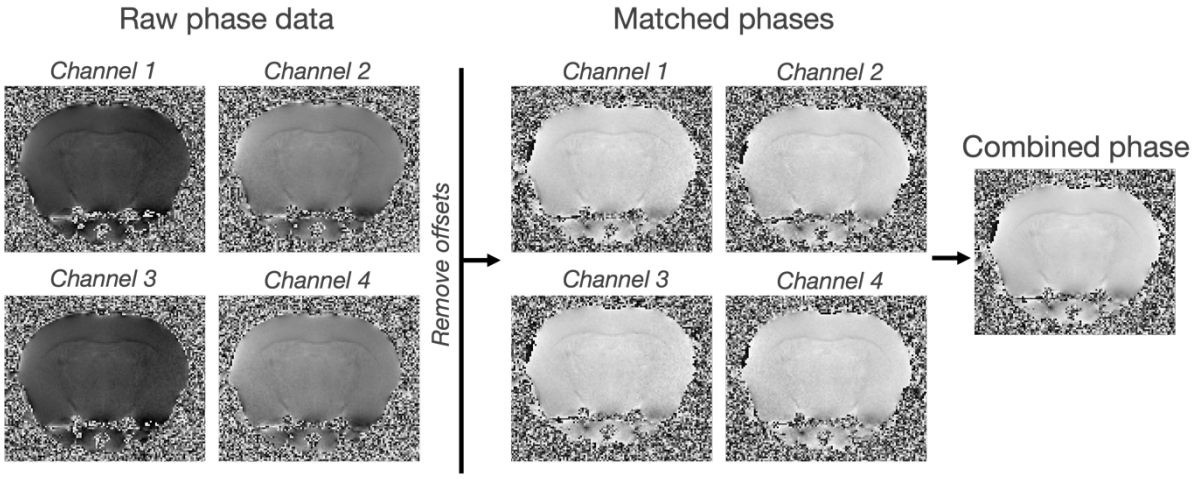

Figure 4.10. Coil combination results in matched phases across the four channels. The resulted combined phase volume is used for subsequent steps in the pipeline.

The field map and a phase offset map are estimated via voxel-wise non-linear fitting of the coil-combined complex data (C. Wang et al., 2020) using custom MATLAB scripts. The field maps from both repetitions are estimated separately at this pipeline stage, averaged prior to the next processing step and then multiplied by a brain mask obtained using an atlas-based segmentation algorithm (Chakravarty *et al.*, 2013). The phase offset map is used to identify whether there are any large, focal variations in phase occurring between the brain and the skull, possibly due to the presence of air bubbles. We empirically derived a threshold to remove any unreliable voxels.

The estimated field maps are affected by the presence of external, smooth field perturbations of low frequency, which need to be removed. Our group found the variable-kernel Sophisticated Harmonic Artifact Reduction for Phase Data (vSHARP) method (Li, Wu and Liu, 2011; Schweser *et al.*, 2011) to yield the most reliable results compared to other widely-used background field removal methods.

Dipole inversion converts the magnetic field offset into χ values (**Equation 3.20**). We use the iterative LSQR (iLSQR) method (Li *et al.*, 2015), a very fast method which also eliminates streaking artifacts, which arise due to ill-conditioning. iLSQR requires the direction of the B_0 field as one of the inputs. Our samples are positioned in the holders

such that the anatomical AP axis is aligned with B_0. We position our field of view accordingly so that the image AP axis is also roughly aligned with the anatomical AP axis.

Finally, because susceptibility is a relative measure, we reference the measured χ values to the whole brain by subtracting the mean susceptibility across the brain from each individual voxel. Our QSM and R_2^* pipelines are depicted in **Figure 4.11**.

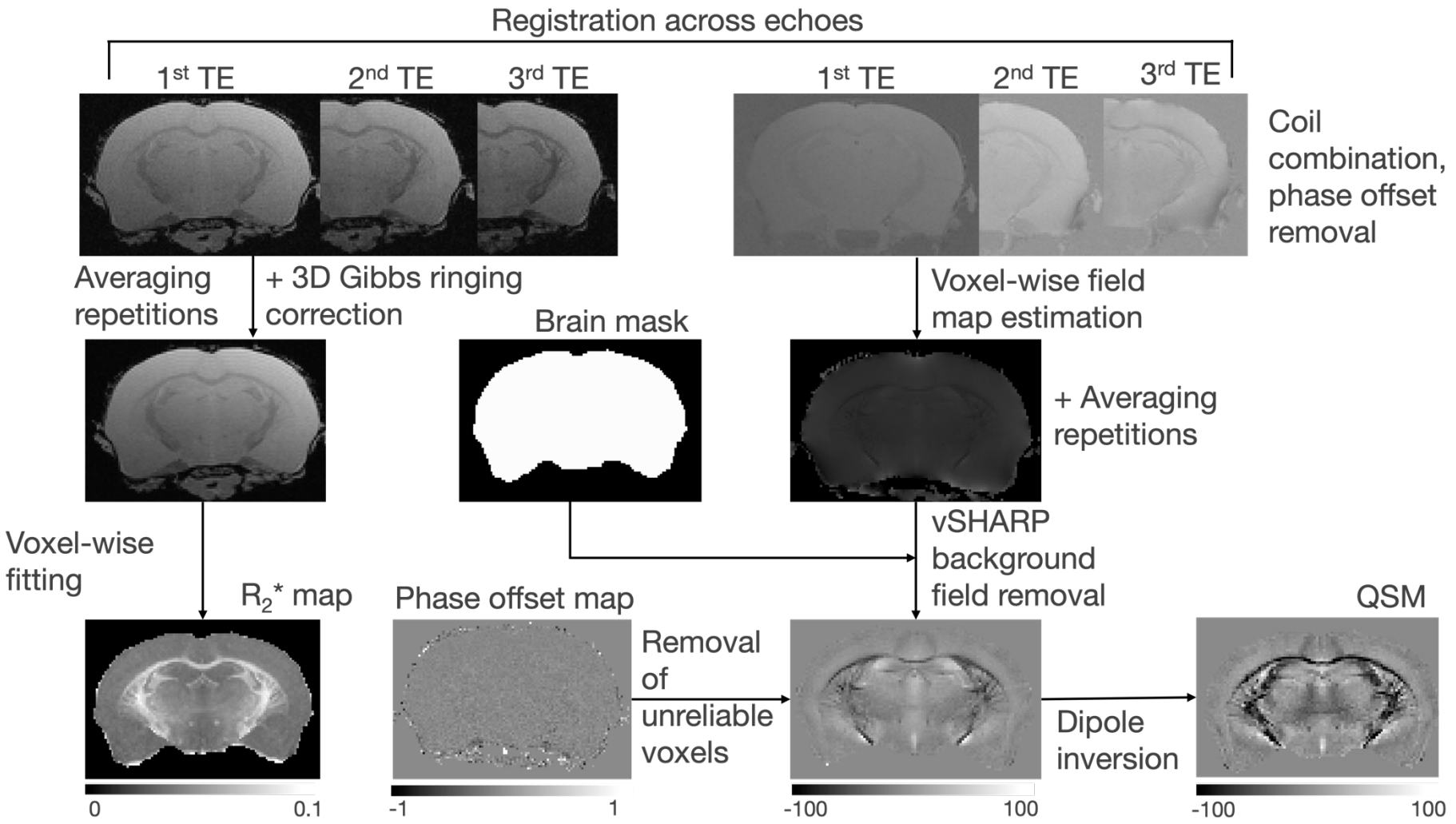

Figure 4.11. Diagram of the QSM processing pipeline, adapted from Wang and colleagues (C. Wang et al., 2020). TE: echo time; QSM: Quantitative Susceptibility Map.

4.5. Co-registration

To facilitate data analysis, particularly registration to histology, the multi-modal MRI datasets need to be co-registered. Slight misalignment due to signal drift can occur within a single acquisition. Furthermore, some of the data had to be separately reacquired after a time interval for some of our samples due to the presence of artifacts. We therefore co-registered all our datasets for each individual sample.

We registered our dMRI to the structural T_2-weighted data using 6 degrees-of-freedom (6DOF) FLIRT registration and the mutual information cost function. The same strategy was used to register the first 10 averaged TEs from the multi-echo MGE data to the first b=0 volume of the diffusion data. Here, we observed an improvement in the registration if an affine transformation (12DOF) was chosen instead of a rigid one (6DOF) for most samples. This can be explained considering that some brain deformations can occur during the storage of the samples, especially given that some of the re-scans were done after a several month-long time interval.

4.6. Quality control

Quality control (QC) for our data acquisition and processing is conducted manually for each individual sample and included as a guide on our pipelines' gitlab pages.

Raw datasets are checked for completeness using FSLeyes. The signal across each volume is compared to previously-acquired scans to evaluate signal homogeneity and artifacts. Specific checks are made for signal stability in the b=0 scans of the dMRI protocol and the signal decay of the multi-echo GRE data to ensure it is roughly exponential. The complex reconstructions of the multi-echo GRE data are also checked.

Our post-processing pipelines output intermediary files at every stage of the processing (as shown in **Figures 4.9** and **4.11**), which are inspected for each individual sample.

4.7. Data open-access

To ensure replicability of results, we plan on releasing the multi-modal MRI data acquired for our cohorts on the Digital Brain Bank (Tendler *et al.*, 2022), our centre's open data release platform for post-mortem imaging datasets. Users will be able to access this data after signing a material transfer agreement.

4.8. Discussion

Working on a new Bruker 7T platform, we developed an *ex-vivo* scanning setup and procedure, including custom-made holders for the brains and 3D printed custom components that would hold the falcon tube in place during scanning, minimising movement due to the reverberations of the gradient coils. The individual scanning sequences were optimised to maximise SNR and tissue contrast, as well as reduce the presence of setup-specific artifacts.

One limitation of our acquisition protocols is the long acquisition time (14h per sample). In particular, the dMRI scan can be shortened in the future by using parallel imaging, partial Fourier or simultaneous multi slice/multiband imaging, while avoiding additional artifacts. As the overnight slot is otherwise unoccupied during the 14h after working hours, additional volumes could be acquired during that time, such as higher-shell b-values that may be used for fitting grey matter microstructural models (Palombo *et al.*, 2020).

Our processing pipelines account for formats and data structures of the acquired datasets and follow established steps to extract the metrics of interest. They were built according to two principles: Firstly, we aimed to minimise the presence of artifacts characteristic to our hardware setup. Secondly, we wanted to extract the metrics of interest using established workflows.

To this day, there is still a lack of consensus regarding processing pipelines for dMRI data (Tax *et al.*, 2022), which may bias research findings and impact reproducibility of results. While the development of processing pipelines for dMRI data is clearly a highly-debated topic, QSM, due to its relative novelty, is still at the stage where such reviews are missing from literature. Further, the code for some of the most popular tools used for QSM processing (e.g., STISuite) is not openly available, making tracking down artifacts while using these tools challenging. Importantly, while it is desirable to standardise processing steps for microstructural MRI pipelines, this may ultimately be an unrealistic goal due to the sheer variability of dataset formats, artifacts and acquisition parameters across research laboratories and scanners.

4.9. Conclusion

In this chapter, we presented the development of multi-modal MRI acquisition protocols and processing pipelines for *ex-vivo* mouse brains. Acquiring *ex-vivo* data enables us to obtain an extensive, high-quality dataset during a 14h overnight acquisition slot. Our data is acquired at high resolution and presents excellent SNR, grey-white matter contrast, being optimal for a range of analyses and for the phenotypic characterisation of novel mouse models.

Chapter 5: Histology in mouse brains prepared for *ex-vivo* MRI: An optimised framework for data acquisition and processing

Context

Quantitative comparisons between MRI and histology can help circumvent MRI's inherent lack of specificity, pinpointing the molecular mechanisms MRI is sensitive to. To achieve this, we need high-quality MRI-histology registration, as well as uniform immunostaining for the extraction of quantitative histological metrics.

One key challenge in performing formalin-fixed, paraffin-embedded (FFPE) immunohistochemistry (IHC) histology experiments in mouse brains prepared for *ex-vivo* MRI is that the tissue suffers molecular alterations due to long storage times, which affects virtually all aspects of the histological processing pipeline. While non-linear image registration tools can deal with deformation of tissue, stained histological slices frequently exhibit more extreme morphological changes like folds, tears, and holes. Further, to extract reliable quantitative metrics, we require uniform staining across each tissue slice, with minimal artifacts.

This chapter presents our development of a histology protocol for mouse brain samples prepared for *ex-vivo* MRI where we address these challenges. Our protocol yields significantly improved results compared to standard histology protocols and satisfies the goals of our MRI-histology framework, allowing for straightforward successful registration to MRI of each individual slice and extraction of quantitative metrics.

Here, we detail the FFPE-IHC protocol optimisations in the first half of the chapter and the optimisation of histological data processing and MRI-histology fusion in the second half.

Published work

Personal contributions

Conference proceedings

Cristiana Tisca, Mohamed Tachrount, Adele Smart, Frederik J Lange, Amy FD Howard, Chaoyue Wang, Lily Qiu, Benjamin Tendler, Claire Bratley, Daniel ZL Kor, Istvan N Huszar, Javier Ballarobre-Barreiro, Marika Fava, Manuel Mayr, Jason P Lerch, Aurea B Martins-Bach, Karla L Miller. *Linking MRI to histology in the mouse brain: A framework for data acquisition and pre-processing.* OHBM Annual Meeting (2023).

Collaborations

Articles

Adele Smart**, Cristiana Tisca**, Istvan N Huszar, Olaf Ansorge, Karla L Miller, Aurea B Martins-Bach. *An optimised protocol for paraffin-embedding and immunohistochemical staining of mouse brain tissue following ex-vivo MRI and prolonged storage.* STAR Protocols (2023).

Conference proceedings

Adele Smart, **Cristiana Tisca,** Istvan N Huszar, Olaf Ansorge, Karla L Miller, Aurea B Martins-Bach. *An optimised protocol for multi-stain, high-throughput histology in mouse brains after ex-vivo MRI.* OHBM Annual Meeting (2023).

Individual contributions to this thesis chapter

My personal contributions to this chapter include: histology experimental design; sample dissection, processing, embedding and slicing; slicing optimisation; optimisation of the versican and brevican IHC staining protocols; quality-controlling and curating the acquired datasets; optimising the MRI-histology registration and co-registration of the MRI-histology datasets for each individual sample; SAF pipeline optimisation for each individual stain; MRI-histology data fusion.

The protocols for processing and embedding samples and for myelin proteolipid protein (PLP) and neurofilament (NF) IHC staining used in this chapter were optimised by Aurea Martins-Bach and Adele Smart. Tissue slicing, IHC staining laboratory work and data digitisation were conducted by Adele Smart. Istvan Huszar and Daniel Kor advised on using and optimising their tensor image registration library (TIRL) and stain area fraction (SAF) pipelines, respectively. The contribution of other collaborators is listed in the **Acknowledgements** section.

5.1. Introduction

MRI modalities are sensitive to numerous characteristics of tissue in the living brain, from fibre orientation and density, myelination and axonal damage to cellularity, iron, and calcium content. This, however, comes with an inherent lack in specificity, coupled with relatively low resolution compared to the size of the tissue components of interest. One way to overcome this drawback is by acquiring MRI and histological stains in the same mouse brains and linking quantitative metrics extracted from histology to MRI metrics. Using immunostains, we can map the tissue distribution of almost any protein of interest and correlate this with the desired MRI metric at the MRI voxel level.

Furthermore, an additional challenge is that histological processing protocols vary considerably across studies and may not specify essential information when described in papers, such as providing examples of full-faced stained tissue slices. They are also regularly not optimised to the long storage times of brain samples prepared for *ex-vivo* MRI.

Provided that optimised histological tissue processing and staining protocols are used, registering the histology data to MRI is done using state-of-the-art registration tools (Ferrante and Paragios, 2017). The TIRL tool (Huszar *et al.*, 2023) facilitates the registration of sparse 2D histological data to 3D MRI data acquired in the same brain, as well as 2D to 2D histology data registration of neighbouring brain slices. TIRL has been successfully used in the MRI-registration of human (Huszar *et al.*, 2023) and primate brains (Howard *et al.*, 2023)

Additionally, most IHC-MRI correlation studies are limited by a low number of samples and data points, mainly due to the manual segmentation of the histological ROIs on very small sections of tissue. The extraction of quantitative histological metrics is also limited by manual, subjective adjustments. Automating ROI segmentation, which can only be done if the registration problem is solved, coupled with the automated extraction of quantitative metrics, can greatly improve results and reproducibility. ROI segmentation can be automated by resampling masks generated via atlas-based segmentation on MRI images into the histology native space, using the

transformations derived from MRI-histology registration. Quantitative histological metrics can be extracted using automated pipelines. An example of such a pipeline was developed in our group (Kor et al., 2022) and provides a quantitative estimate of the fraction of the chromogen, 3,3'-Diaminobenzidine (DAB)-stained pixels within a given patch. This pipeline also requires minimal adaptations to work with large multi-stain histology datasets.

We designed an FFPE-IHC protocol which yields significantly improved histology data quality in *ex-vivo* mouse brain samples prepared for MRI. Further, we optimised our staining protocols for myelin PLP, NF, versican and brevican. The staininig optimisation took about 7 months.

Following data acquisition using these optimised protocols, we adapted TIRL (Huszar *et al.*, 2023) to provide a robust MRI-histology registration, accounting for the high resolution of this type of data. Finally, we conducted a per-stain optimisation of an automated stain area fraction pipeline (Kor et al., 2022) for the myelin PLP, NF, brevican and versican stains. Optimising these tools for our datasets took about 5 months.

After the optimisation of the protocols, we acquired and processed multi-stain histology data in our two cohorts, *Vcan* and *Bcan*. The data acquisition took about 2 months. The cohorts and study results will be presented in **Chapter 6**.

5.2. Experimental design considerations

5.2.1. Study hypotheses

The mouse models we investigate in this thesis have modifications in the *Vcan* and *Bcan* genes. Versican and brevican are the proteins produced by the genes modified in our mouse models. As described in **Chapter 2**, these proteins interact with both

myelin and neurites. Versican and brevican are abundant during white matter development and act as myelin inhibitors. Similarly, they also inhibit neurite outgrowth in the mature brain. Therefore, we hypothesise that the absence or change in versican or brevican could be reflected in the content of myelin or neurites, and that this would drive MRI signal changes in the mouse models.

5.2.2. Selection of immunostains

To test our hypothesis, we acquired multiple stains in each individual mouse brain: myelin, neurofilaments (NF), and either versican (for the *Vcan* mouse model) or brevican (for the *Bcan* mouse model).

Myelin is thought to be a major determinant of signal variation in MRI and is of broad interest in both clinical and basic neuroscience. IHC myelin stains that target specific myelin proteins (e.g., PLP) show strong, continuous staining alongside axonal fibres, providing excellent grey-white matter contrast (**Figure 5.1** a, red arrow). Myelin is also a major determinant of contrast in a range of MRI neuroimaging modalities. As described in **Chapter 3**, myelin has negative magnetic susceptibility and is a major source of image contrast in QSM. Lipids and proteins in the myelin sheath also influence the T_1, T_2 and T_2^* relaxation rates. Finally, myelin also hinders the diffusion of water perpendicular to the axonal direction, which is reflected by dMRI metrics extracted from a wide range of biophysical models.

Neurofilaments (NF) are cytoskeletal protein chains found in neuronal dendrites and axons. NF stains give insight into the quantity and organisation of neurites. NF IHC stains are dense and ubiquitous, with reduced grey-white contrast compared to myelin stains (**Figure 5.1** b). They are continuous in fibres parallel to the slicing plane (**Figure 5.1** b, red arrow) and appear punctate when viewed cross-sectionally in fibres perpendicular to the slicing plane (**Figure 5.1.** b, green arrow).

Versican and brevican protein distribution in histological tissue has never before been qualitatively or quantitatively linked to MRI metrics. The versican isoforms have

different staining patterns: versican V2 shows strong staining in the white matter, along myelinated fibres, with punctate staining accumulations at the nodes of Ranvier (Dours-Zimmermann et al., 2009). Versican V1 has a limited expression in the adult brain, confined to the choroid plexus, meninges and blood vessels (Dours-Zimmermann et al., 2009). Brevican shows more diffuse staining in white matter, with no accumulations at the nodes of Ranvier. Brevican accumulates, however, in perineuronal nets around the neuronal nuclei (Dours-Zimmermann et al., 2009).

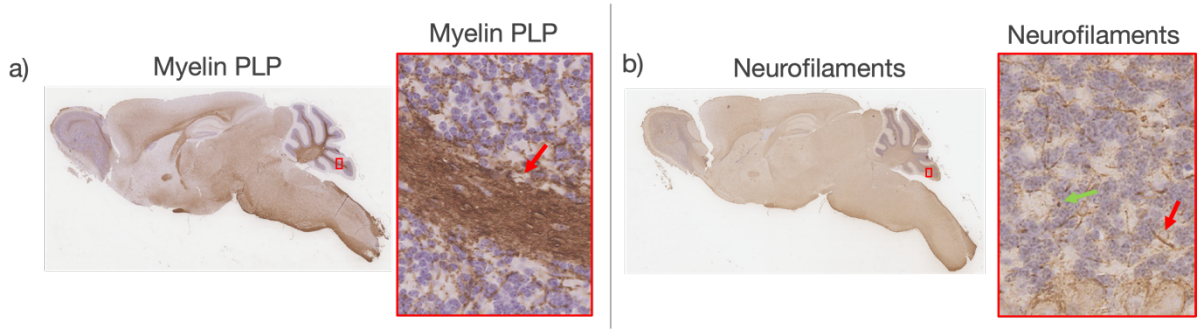

Figure 5.1. Examples of FFPE-IHC stained wild-type mouse brain slices. a) Myelin PLP; b) NF-stained sagittal brain slices.

Versican and brevican immunostains facilitate a complete characterisation of our mouse models of interest. We expect no brevican to be present in our knock-out *Bcan* mice (as detailed in **Chapter 6**). However, the *Vcan* mice are not knock-outs, but present a genetic modification that leads to impaired cleavage of the versican V1 isoform (as detailed in **Chapter 6**). The effect of this genetic modification on the distribution of the versican isoforms in the brain has not been investigated in this model. Therefore, to elucidate this question, we wanted to acquire IHC stains for both versican V1 and V2 in the *Vcan* mice.

5.2.3. Data acquisition

Our MRI and histology analyses focus on the major white matter tracts in the brains of our mouse models of interest: the corpus callosum, fimbria, internal capsule and

anterior commissure. We therefore acquired histological data at three different locations (A, B and C) spanning the mouse brain. These locations include sections through all the white matter tracts we are interested in. We opted for coronal brain slices. Coronal slices are more compact compared to other planes (axial or sagittal) and their morphology is more easily preserved during slicing, which aids MRI-histology registration.

2 samples per group – male wild-type, female wild-type, male homozygous and female homozygous – were selected from each cohort (*Vcan* and *Bcan*). The number of samples was limited by the time available for the thesis research.

We chose FFPE-IHC due to its excellent conservation of brain morphology and ease of processing and long-term storage, coupled with our group's extensive expertise in acquiring and processing this type of data. Further, our group has developed optimised protocols for myelin PLP and NF IHC staining of mouse brains prepared for *ex-vivo* MRI.

We selected 4 consecutive slides for each *Vcan* mouse brain sample (for myelin PLP, NF, versican V1 and versican V2 staining) and 3 consecutive slides for the *Bcan* mouse brain samples (for myelin PLP, NF and brevican staining) at each location of interest (A, B, C, **Figure 5.2**). The optimised histological slice thickness was 4μm. Compared to the 60μm voxel size of the T_2-weighted structural data, the 3-4 slices at each location of interest should all be in the same MRI voxel. This estimate holds even after accounting for the tissue shrinkage that occurs during tissue processing and paraffinization (Rodgers *et al.*, 2022). During the IHC staining experiments, we occasionally got random artifacts that could affect stain quantification in the regions of interest. In such cases, the slides need to be re-stained: the closest slice was chosen which did not present any tears or folds.

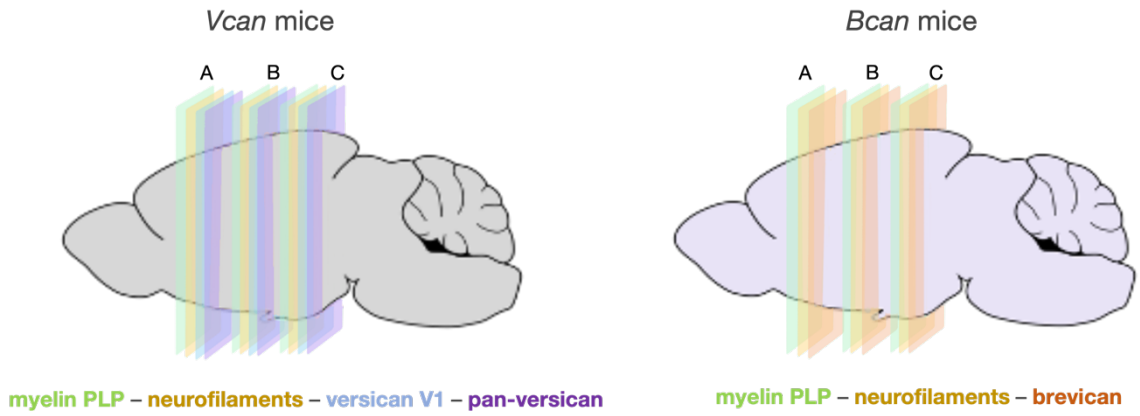

Figure 5.2. Histology slice locations (A, B, C) for the *Vcan* and *Bcan* mice. Figure produced with help from Adele Smart.

5.3. Histological data acquisition: methodology and results

The generic histological protocol that we optimised is presented in **Chapter 3**, **Figure 3.7**. Here, we detail our contributions to each individual step. The step-by-step protocol and a comprehensive list of the reagents, solutions and equipment used are detailed in the co-written protocol paper (Smart *et al.*, 2023). In the optimisation stage of this project, we worked with surplus mouse brain samples perfused for MRI. For the stain optimisation for the versican stain, which required the visualisation of staining patterns in other organs, we also used organs from a freshly-perfused wild-type mouse.

5.3.1. Brain dissection

As a first step, the brains need to be removed from the skull (**Figure 5.3** A-D). Care must be taken to avoid damaging the tissue. Any tear is amplified in the subsequent steps of the pipeline and can result in a full break during slicing. We recommend doing this slowly with small tweezers and scissors.

5.3.2. Brain cutting

We used a brain slicing matrix (Mouse Coronal 40-75g, Electron Microscopy Sciences) for adult mouse brains to cut the brain using a razor blade slid through notches 4 and 12 (Bregma +3 and -5) (**Figure 5.3** H). We removed the olfactory bulb and cerebellum during sectioning. The remaining brain slab contains the locations of interest shown in **Figure 5.2**.

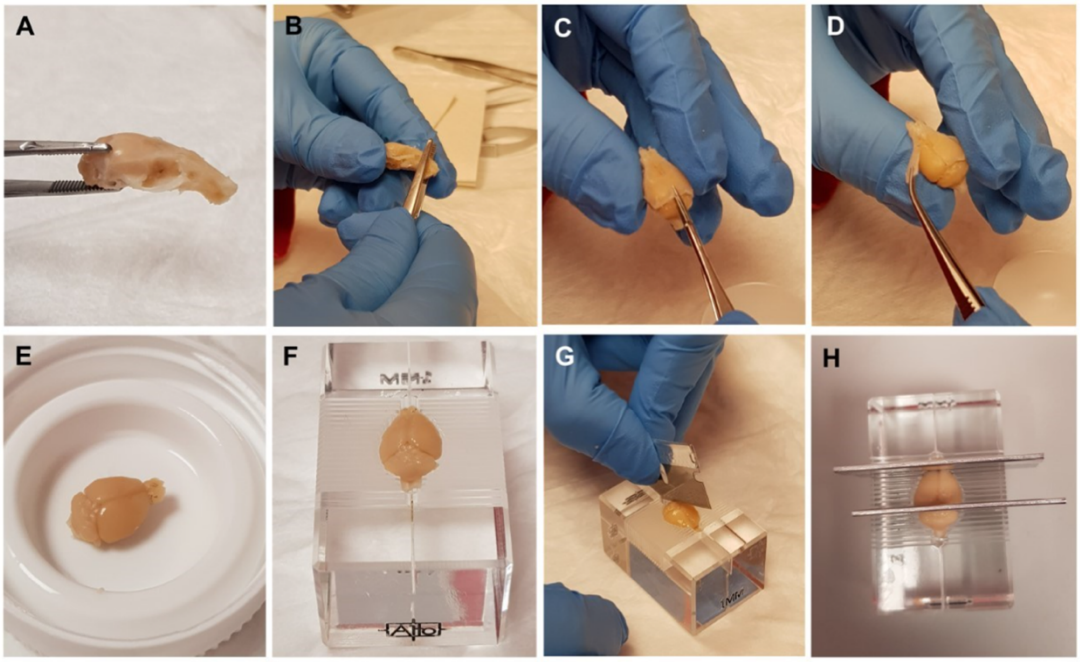

Figure 5.3. Presentation of our dissection and brain cutting process. A. Example mouse brain in the skull. B-D. Removal of the brain from the skull. E. Dissection outcome. F. Mouse brain in the brain slicing matrix. G. Example sagittal cutting of the brain. H. Optimised coronal cutting of the brain.

5.3.3. Tissue processing and embedding

The FFPE-IHC processing stage includes dehydration, clearing and wax infiltration. This protocol was optimised for small mouse brain sections from samples stored in PBS, sodium azide and Gd-based contrast agent solution for long periods of time.

Our samples were stored for 27 months before processing, primarily due to facility closures during the COVID-19 pandemic. The optimised processing protocol is presented in **Table 5.1**. The protocol was optimised by trial and error starting from a protocol used for fresh brain samples. Notably, the final dehydration, clearing and infiltration in wax steps in this routine are shorter than in routines used for fresher samples.

The embedding stage involves manually inserting the paraffin-infiltrated sections into wax blocks using an embedder. We embedded a single coronal brain section in the middle of each individual block. We positioned the caudal part of the coronal mouse brain section on the bottom of the mould, which is where we began slicing.

Notably, the final dehydration, clearing and infiltration in paraffin wax steps of this routine are shorter than in routines used for fresher samples. A standard, longer routine results in difficulty during slicing, as well as tissue folds and tears. Our improved results are highlighted in **Figure 5.4**.

Step		Duration
Manual stage	Rinse tissue sections in distilled water	2x – 30s each
	70% ethanol	12h
	80% ethanol	45min
	95% ethanol	2x – 45min each
Automatic stage	100% ethanol	3x – 15min each
	100% ethanol	3x – 15min each
	Clearing solution (Citroclear, Genta Medical)	30min
	Clearing solution (Citroclear, Genta Medical)	45min
	Clearing solution (Citroclear, Genta Medical)	1h
	Paraffin Wax (SLS Ltd)	30min
	Paraffin Wax (SLS Ltd)	45min
	Paraffin Wax (SLS Ltd)	1h

Table 5.1. Timings of the optimised processing protocol (Smart *et al.*, 2023).

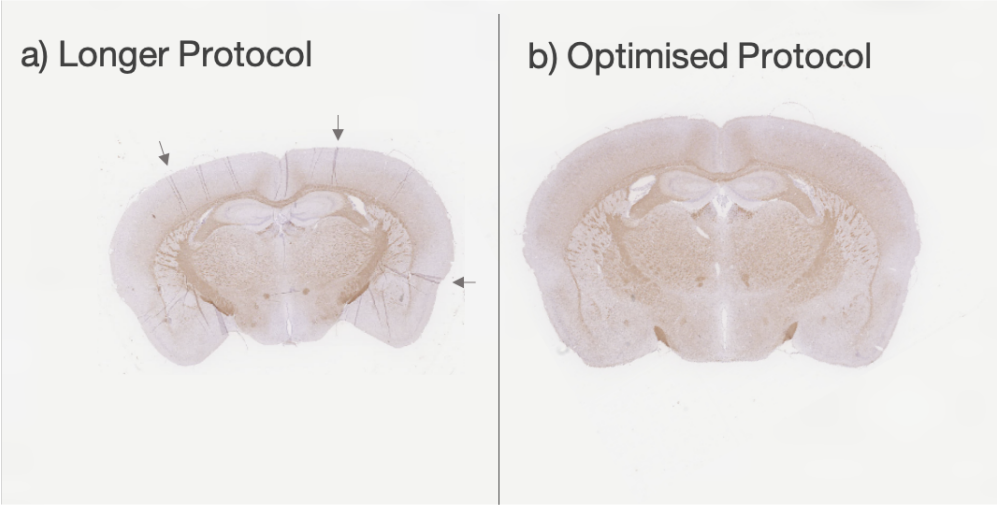

Figure 5.4. Representative images of myelin PLP stained slices from (a) a sample processed with using a longer protocol (1h longer in ethanol 95%), compared to (b) a sample processed using our optimised protocol. Sample (a) shows a notable brain shrinkage, accompanied by tissue folds throughout the slice which change brain morphology. The artifacts would compromise the registration of the slice to the MRI data acquired in the same brain and would confound quantification (e.g., stain area fraction values).

5.3.4. Tissue slicing

A microtome is used to obtain thin slices. The paraffin blocks were trimmed and soaked in a softening agent (Mollifex, VWR) prior to slicing. The block was then soaked in ice water for 5min after removing approximately 20 slices to reduce tears and artifacts. We conducted experiments to identify the optimal slice thickness given the quality of our samples, finding 4μm thick slices to produce the fewest folds, tears or breaks (**Figure 5.5**).

We also designed a strategy for the visual identification of the three locations of interest (A, B, C, **Figure 5.2**). We advanced in increments of 100μm (cutting at 10μm slice thickness) until we reached the proximity of each location of interest, which was inspected under the microscope and compared to atlas references, primarily focusing on the shape of white matter tracts. We then proceeded to obtain around 20 slices at

each location. More slices were obtained if we encountered artifacts such as tissue folds or breaks during water bath floating. To ensure consistency, we aimed to collect brain slices onto the slides in the same orientation. Particular care was taken to avoid flipping of the hemispheres (for instance if the slice was placed in the water bath on the wrong side). We also ensured that the position and orientation of the cutting was maintained across samples to aid in the initialisation of the MRI-histology registration.

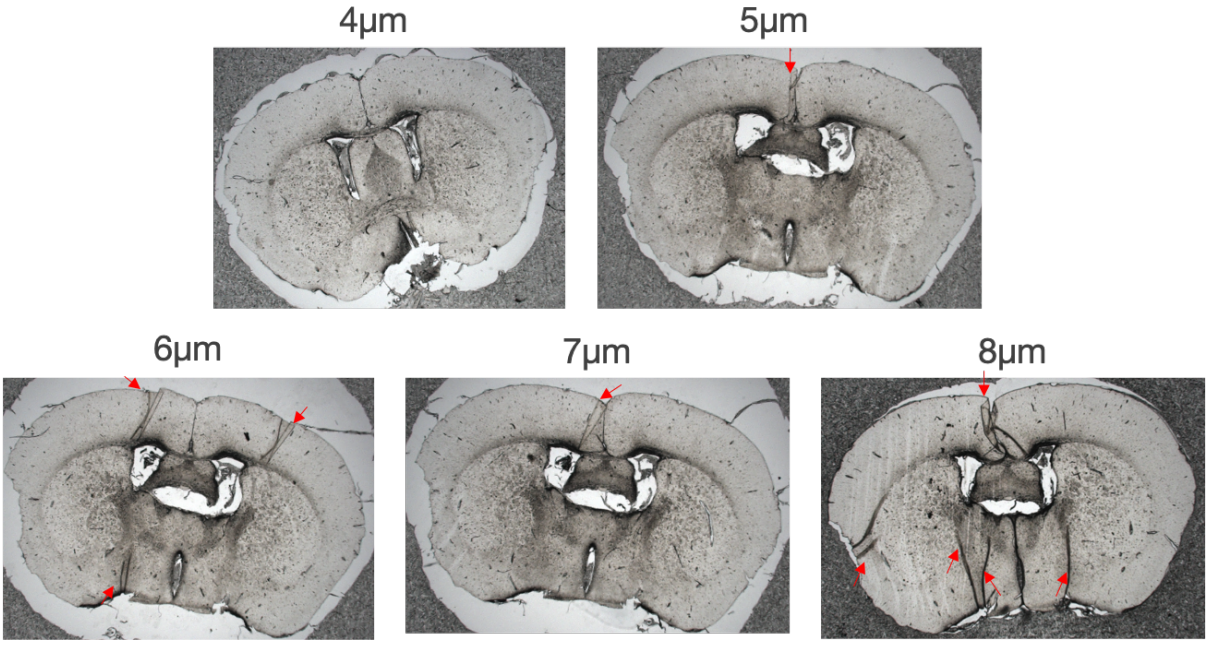

Figure 5.5. Light microscope photographs of coronal sections acquired at different slice thicknesses, between 4μm and 8μm, in a wild-type mouse. The dark background is due to the opacity of paraffin wax, as the slices were still paraffin-embedded. The red arrows show tissue folds which compromise the morphology of the slices. They are accentuated by an increase in slice thickness.

5.3.5. Immunohistochemistry

When choosing a particular manufacturer's antibody, researchers often select one that satisfies their experimental setup (e.g., IHC vs immunofluorescence) and has been extensively validated in literature. Developing an optimised protocol can be a challenge depending on each manufacturer's antibody, as well as on the sample quality. This was the case the anti-versican and anti-brevican antibodies, for which very limited information regarding the recommended protocol was available.

In this sub-section, we detail the protocols for myelin PLP and NF stains and the testing and optimisation we performed for the versican and brevican stains. While all protocols required some adaptation, we found versican and brevican stain optimisation to be particularly challenging.

The IHC protocol steps are presented in **Figure 5.6**. Further details about the optimisation of each step are highlighted in **Appendix B**. Our chosen experimental condition for each step was the one that yielded the stain quality that was closest to the examples from literature (**Figure 5.1**, b). A high-level description of our optimised staining protocols is given in **Table 5.2**.

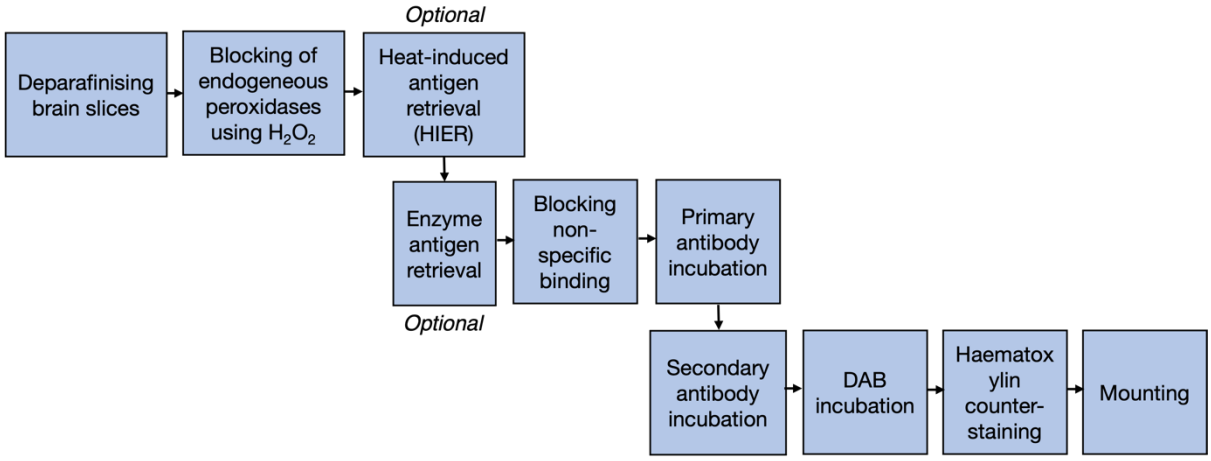

Figure 5.6. Immunohistochemistry staining protocol steps. The steps labelled *'Optional'* are only required for certain targets.

All stained brain slices acquired as part of our studies were digitised in the SVS format using a slide scanner (Leica ScanScope AT Turbo) at x40 magnification, achieving a resolution of 0.25μm/pixel, corresponding to a typical image size of approximately 32,000 x 42,000 pixels at full resolution.

Primary antibody name	Company, code	H_2O_2 incubation time	Heat-induced antigen retrieval	Chondroitinase enzyme antigen retrieval	Blocking buffer solution	Antibody concentration	Primary antibody incubation time
Anti-brevican	Abcam, ab285162	30min	Autoclave in citrate buffer	None	Dako Protein Block Serum Free	1:2000	1h at room temperature
Anti-versican (V0/V1)	Merck, AB1033	1h	None	37°C for 1h	Dako Protein Block Serum Free	1:500	Overnight (18h) at 4°C
Anti-versican (pan-versican)	BioTechne (Novus), NBP1-85432	30min	Autoclave in citrate buffer	37°C for 1h	Dako Protein Block Serum Free	1:500	1h at room temperature
Anti-myelin proteolipid protein (PLP)	Abcam, ab28486, RRID: AB_776593	30min	Microwave in citrate buffer	None	Fetal Calf Serum, 10%	1:500	1h at room temperature
Anti-neurofilaments (NF)	Sigma-Aldrich, N4142, RRID: AB_477272	30min	Autoclave in citrate buffer	None	Fetal Calf Serum, 10%	1:2000	1h at room temperature

Table 5.2. Optimised protocol details for the brevican, versican (V0/V1), pan-versican, myelin PLP and neurofilaments stains.

5.3.5.1. Myelin PLP and neurofilaments immunostaining

The myelin PLP and NF-stained slices were obtained using the protocol described in our publication (Smart *et al.*, 2023). Summary details are provided in **Table 5.2**. The optimisation of these two staining protocols was not done as part of this thesis.

Our *Vcan* and *Bcan* myelin PLP datasets were acquired in two batches. The first staining batch included random artifacts on some slides. Neighbouring slides were manually selected for re-staining in the second batch. Example myelin PLP-stained slices at the three locations of interest in one wild-type mouse are shown in **Figure 5.7**. Our myelin PLP dataset presents the expected contrast and stain specificity, highlighting the complex arrangement of myelinated fibres.

Our *Vcan* and *Bcan* NF datasets were acquired in three batches. This was due to random artifacts on some slides. Neighbouring slides were manually selected for re-staining in the second and third batches. Example NF-stained slices at the three locations of interest in one wild-type mouse are shown in **Figure 5.8**. Our NF data showcases the expected staining pattern and specificity, illustrating the elaborate organisation of neurites in the rodent brain.

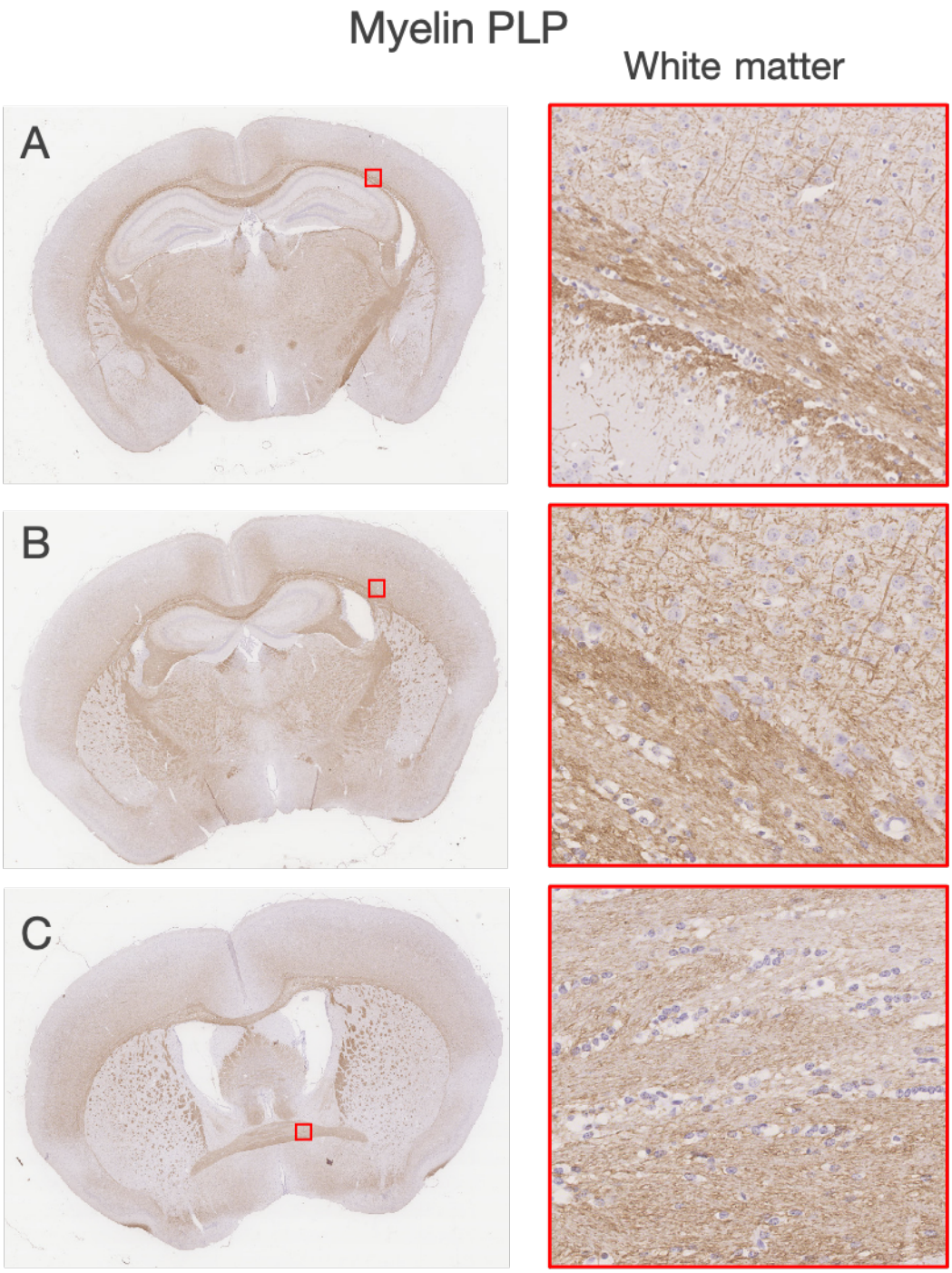

Figure 5.7. Example representative myelin PLP-stained slices (digitised) at the three locations of interest (A, B, C) in a wild-type *Bcan* mouse brain, alongside a zoomed-in insert (red) showing a close-up of staining in white matter.

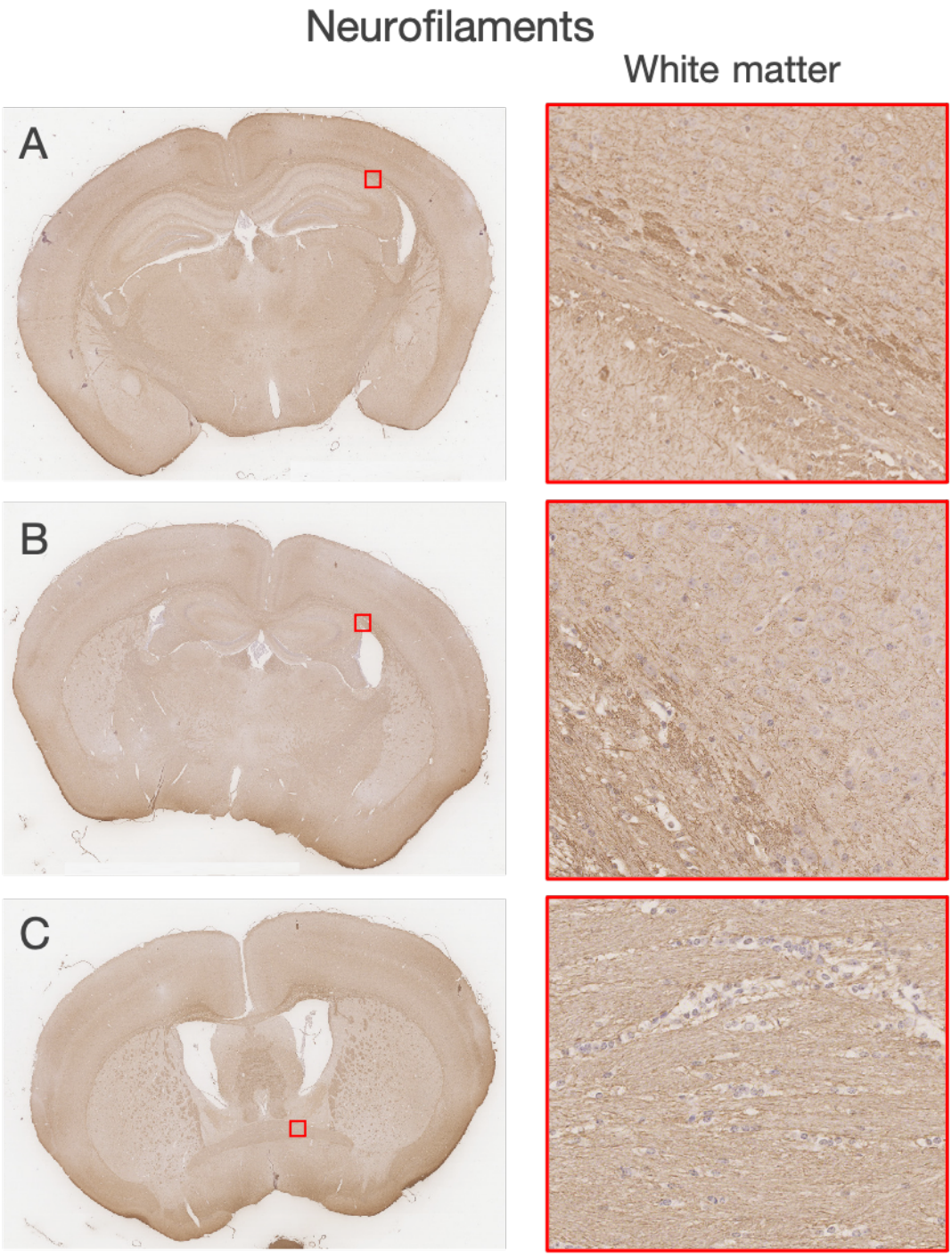

Figure 5.8. Example representative NF-stained slices (digitised) at the three locations of interest (A, B, C) in a wild-type *Bcan* mouse brain, alongside a zoomed-in insert showing a close-up of staining in white matter (red).

5.3.5.2. Optimisation of the versican and brevican immunostains

Following unsuccessful results for our initial antibody choices – chosen due to their extensive use in literature – we decided to test and optimise multiple antibodies from different companies in parallel. We reviewed all remaining available anti-brevican and anti-versican commercial antibodies and selected 2-3 antibodies fitting our experimental criteria.

To identify the best-performing antibody for samples prepared with the protocol described above, we evaluated the results qualitatively based on the following: homogeneity of the stain across the brain slice, intensity of the positive stain, level of non-specific background staining, and, in some cases, the level of staining/lack of staining observed in other organs (e.g., we do not expect brevican staining outside the brain).

The protocol optimisation for the anti-versican and anti-brevican antibodies focused on the H_2O_2 incubation time, testing various HIER methods, chondroitinase ABC (ChABC) enzyme antigen retrieval, blocker type, primary antibody concentration and primary antibody incubation time. We used freshly-prepared solutions and assessed the quality of the reagents prior to each staining.

5.3.5.2.1. Versican

As detailed in **Chapter 2**, versican is present in tissues in four isoforms: V0-V3, with V0, V1 and V3 expressed in all tissues, and V2 expressed only in the brain.

We considered the available commercial anti-versican antibodies. These antibodies target V0 and V1, V0 and V2 and pan-versican (all isoforms). V0 and V3 have extremely limited expression in the brain (Perissinotto *et al.*, 2000; Harten *et al.*, 2020), while V1 has very low expression (Dours-Zimmermann et al., 2009). Therefore, we will refer only to the dominant versican protein isoform (either V1 or V2) for each stain in the following sections.

Versican V1

We successfully optimised the V0 and V1 (Merck, AB1033) antibody. We evaluated the stain specificity in non-brain organs, where expression of V1 is much higher. The strongest expected staining was in the heart's aortic valve (Stephens, Chu and Grande-Allen, 2008), which we used as a positive control (**Figure 5.9**). We also evaluated staining patterns in the kidney, cartilage and lungs. We selected a concentration that had the lowest non-specific background staining in the brain, yet still revealed specific staining in the heart's aortic valve, choroid plexus and vessel walls. We found heat retrieval to increase non-specific staining in the brain and therefore did not include this step in the final protocol. Overnight primary antibody incubation, a longer H_2O_2 blocking step and the Dako Protein Block Serum free were selected due to improved staining specificity.

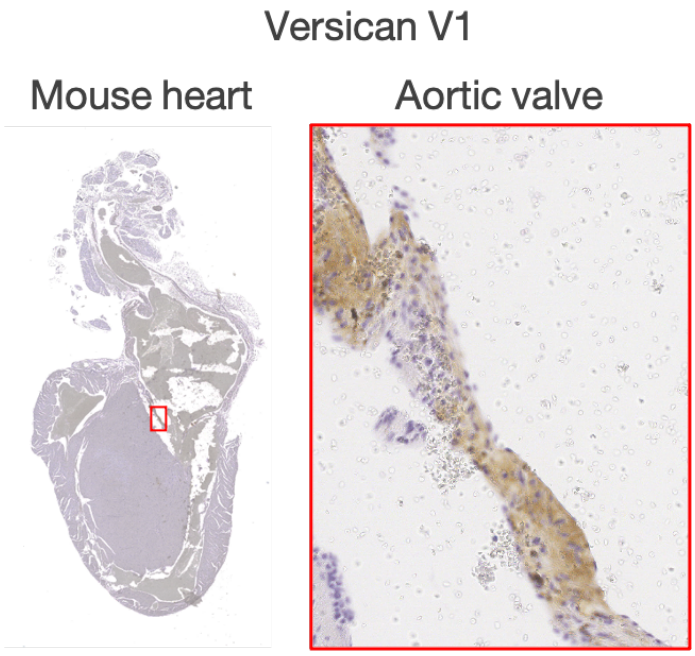

Figure 5.9. Example versican V1-stained wild-type mouse heart, stained using our optimised V1 protocol. We observe specific staining throughout the heart, particularly the expected strong staining in the aortic valve.

The optimised protocol details are in **Table 5.2**. Our versican V1 datasets for the *Vcan* wild-type and homozygous mice were acquired in three batches. This was due to random non-uniform staining present on some of the stained brain slices. Example versican V1-stained brain slices at the three locations of interest in one wild-type mouse are shown in **Figure 5.10**.

We observed the expected lack of staining in the adult rodent brain, with V1 staining confined to the choroid plexus, as well as some vessel walls. Intriguingly, we detected specific staining in some areas of the mouse brain. This, to our knowledge, has not been studied before or mentioned in literature. We observed punctate staining around cell bodies in the CA2 area of the hippocampus and latero-dorsal thalamic nucleus and an accumulation of positive staining in the cell bodies of the median preoptic nucleus. This observation held for all 8 mice we performed histology experiments on (4 wild-types and 4 *Vcan* homozygous mice). Although it is yet to be described, this staining pattern is present in the Allen Brain Atlas in-situ hybridisation data as well (https://mouse.brain-map.org/experiment/show?id=73520975). This data targets RNA molecules instead of proteins within a histological section.

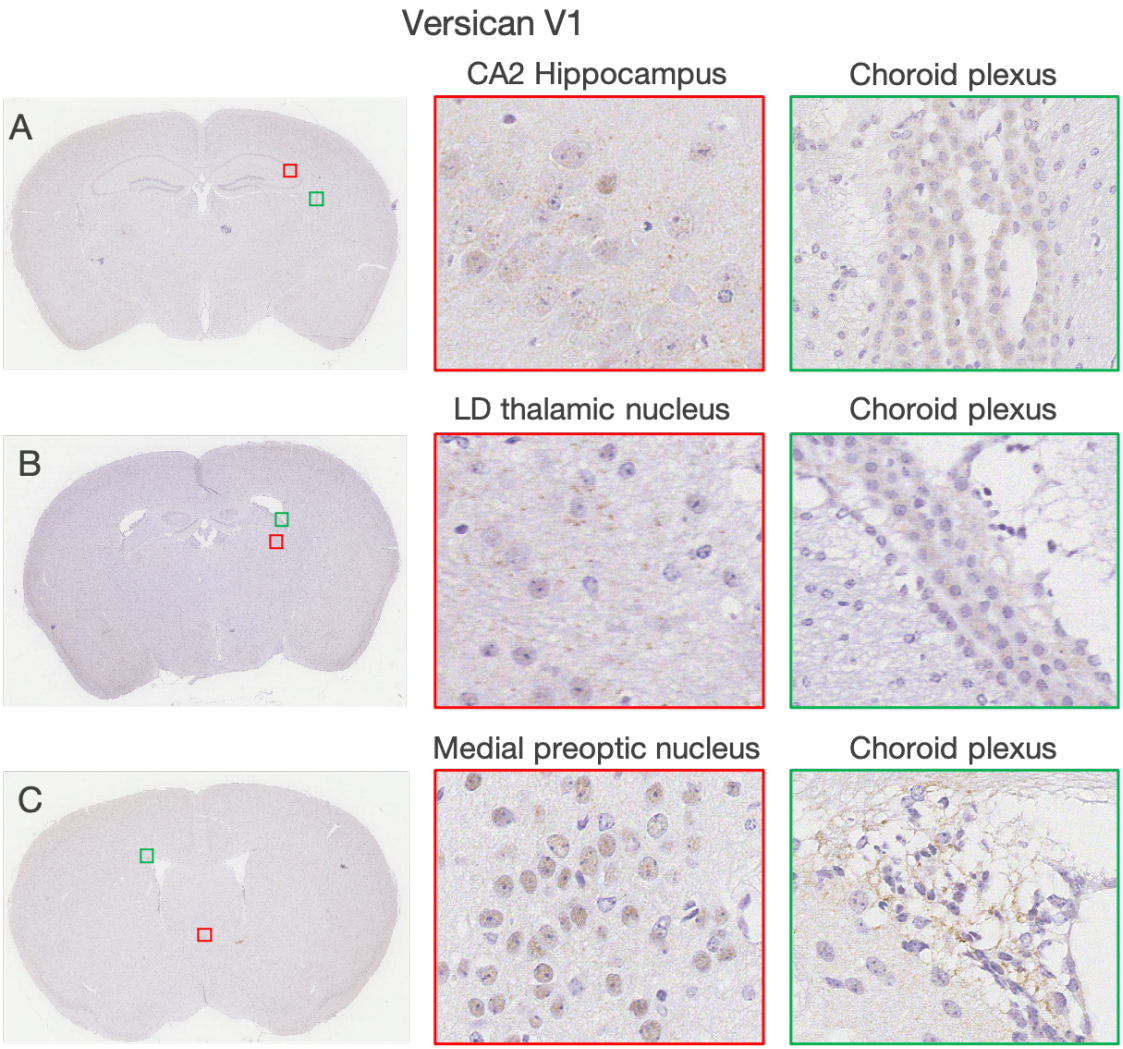

Figure 5.10. Example representative versican V1-stained brain slices (AB1033, 1:500, digitised) at the three locations of interest (A, B, C), in wild-type *Vcan* mouse brains, alongside zoomed-in inserts showing positive, specific staining in various brain areas (red, red arrows) and in the choroid plexus (green, green arrows). LD: latero-dorsal.

Versican V2

V2 is the dominant isoform of versican in the brain. Thus, antibodies characterised as V0/V2 and pan-versican (targeting all versican isoforms) are essentially V2 stains, although the latter may have some staining due to the V1 isoform. Here, we consider both categories of antibodies.

We initially selected a versican V0 and V2 antibody (AB1032, Merck), due to being more specific to the V2 isoform, as well as due to its widespread use in literature (Zhou *et al.*, 2016; Ueno *et al.*, 2017; Evanko *et al.*, 2018; Foulcer, Day and Apte, 2022). However, our optimisation experiments failed, despite numerous adjustments, including testing multiple antibody batches. We obtained variable staining results, with severe artifacts across brain slices even when repeating experiments while strictly controlling all experimental variables. The stained brain slices show strong staining near the edges of the brain and faint staining throughout the slide (**Figure 5.11**, a), which contrasts with the expected result: strong staining in the white matter (Dours-Zimmermann et al., 2009).

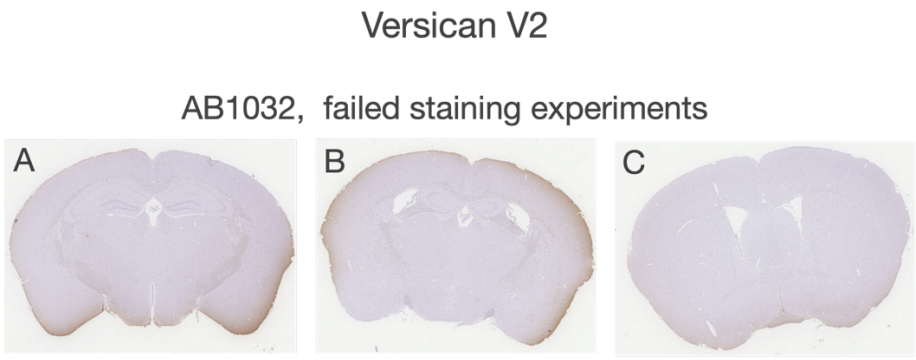

Figure 5.11. Representative failed staining experiments using the anti-versican Merck AB1032 antibody at the three locations of interest (A, B, C), in a wild-type mouse brain from the *Vcan* cohort. Experimental details: Anti-versican V2 Merck AB1032 antibody, 1:150 concentration, overnight primary antibody incubation, HIER: autoclave citrate, chondroitinase enzyme antigen retrieval, Dako Protein Block Serum Free.

All other available and validated commercial antibodies were pan-versican antibodies. We subsequently tested pan-versican antibodies from Novus/BioTechne (NBP1-85432) and Abcam (ab177480). The antibody from Novus/BioTechne outperformed the Abcam antibody in our experiments. A standard stain optimisation schedule enabled us to narrow down on the optimal protocol parameters (**Table 5.2**). Example results are shown in **Figure 5.12**. These results exhibit the expected grey-white matter

contrast, with increased versican V2 staining in the white matter and punctate staining patterns at the nodes of Ranvier. The slides had minimal staining artifacts, which enabled us to extract reliable quantitative staining metrics from this dataset. The Novus/BioTechne antibody also showed consistent results across experiment replications. Our *Vcan* pan-versican dataset was acquired in two batches.

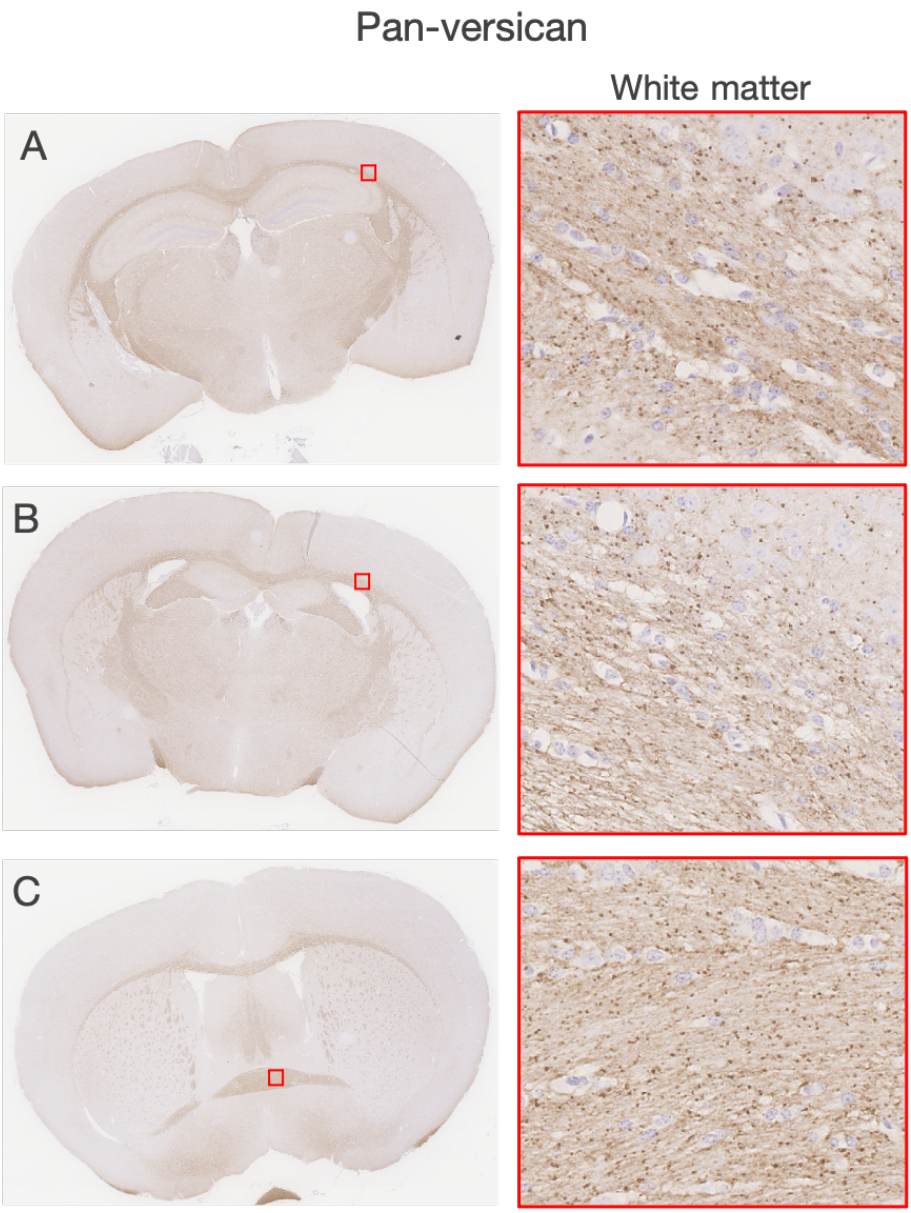

Figure 5.12. Example representative pan-versican stained slices (Novus NBP1-85432, 1:500), digitised at the three locations of interest (A, B, C), in a wild-type mouse brain from the *Vcan* cohort, alongside zoomed-in inserts showing positive, specific staining, including punctate staining in white matter, indicating the versican V2 accumulations at the nodes of Ranvier.

5.3.5.2.2. Brevican

There are limited FFPE-IHC examples of brevican stains in literature, including no example of a full-faced brain slice. We identified an example full-faced coronal mouse brain slice stained against brevican using immunofluorescence (https://www.proteinatlas.org/ENSG00000132692-BCAN/brain). This provides a reference for the expected grey-white matter contrast. Here, we observed an inverse contrast compared to the versican stain: there is increased staining in the grey matter compared to white matter.

We therefore expected to observe the same contrast as in **Figure 5.13**, b in our stained brain slices. Furthermore, when optimising our brevican stain, we sought to achieve the characteristic staining pattern in perineuronal nets, surrounding neuronal cell bodies, some degree of diffuse staining in white matter as described in literature (Dours-Zimmermann et al., 2009) and, importantly, no brevican staining outside the brain or in the brain of knock-out mice. Failed brevican immunostaining experiments using the Novus NBP2-15616 antibody showed non-specific staining outside the brain and in *Bcan* knockouts (which lack brevican) (**Figure 5.13**, a). Following this, we tested multiple antibodies simultaneously.

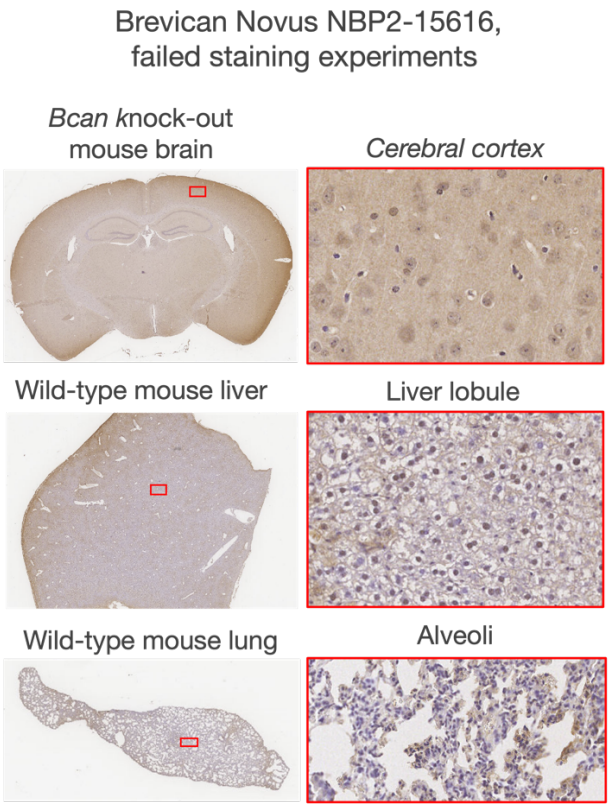

Figure 5.13. a) Representative failed staining experiments using the anti-brevican Novus NBP2-15616 antibody. We observe false-positive staining in a *Bcan* knock-out mouse brain and in the liver and lung of a wild-type mouse (right side of the Figure). No staining is expected in these tissues. Experimental details: 1:100 concentration, overnight primary antibody incubation, HIER: autoclave citrate, chondroitinase enzyme antigen retrieval, Dako Protein Block Serum Free.

Anti-brevican antibodies from ThermoFisher (19017-1-AP), Sigma-Aldrich (HPA007865) and Abcam (ab285162) were tested. Only Abcam (ab285162) showed negative staining outside the brain and in knock-out mice. Following this, we optimised the Abcam antibody to minimise background staining, achieve the expected grey-white matter contrast and specific staining. The details of the optimised protocol are found in **Table 5.2**. Example staining results in a wild-type mouse are highlighted in **Figure 5.14**. We observe the same staining pattern as the one depicted in our reference immunofluorescence image (**Figure 5.13,** b), with reduced staining in white matter compared to grey matter. Diffuse staining is present in white matter as well

(**Figure 5.14**, red insert). We also observe characteristic brevican accumulations in perineuronal nets (**Figure 5.14**, green insert, black arrows). Importantly, no staining whatsoever is present in *Bcan* knock-out mice (**Figure 5.15**). Our *Bcan* datasets were stained in a single batch.

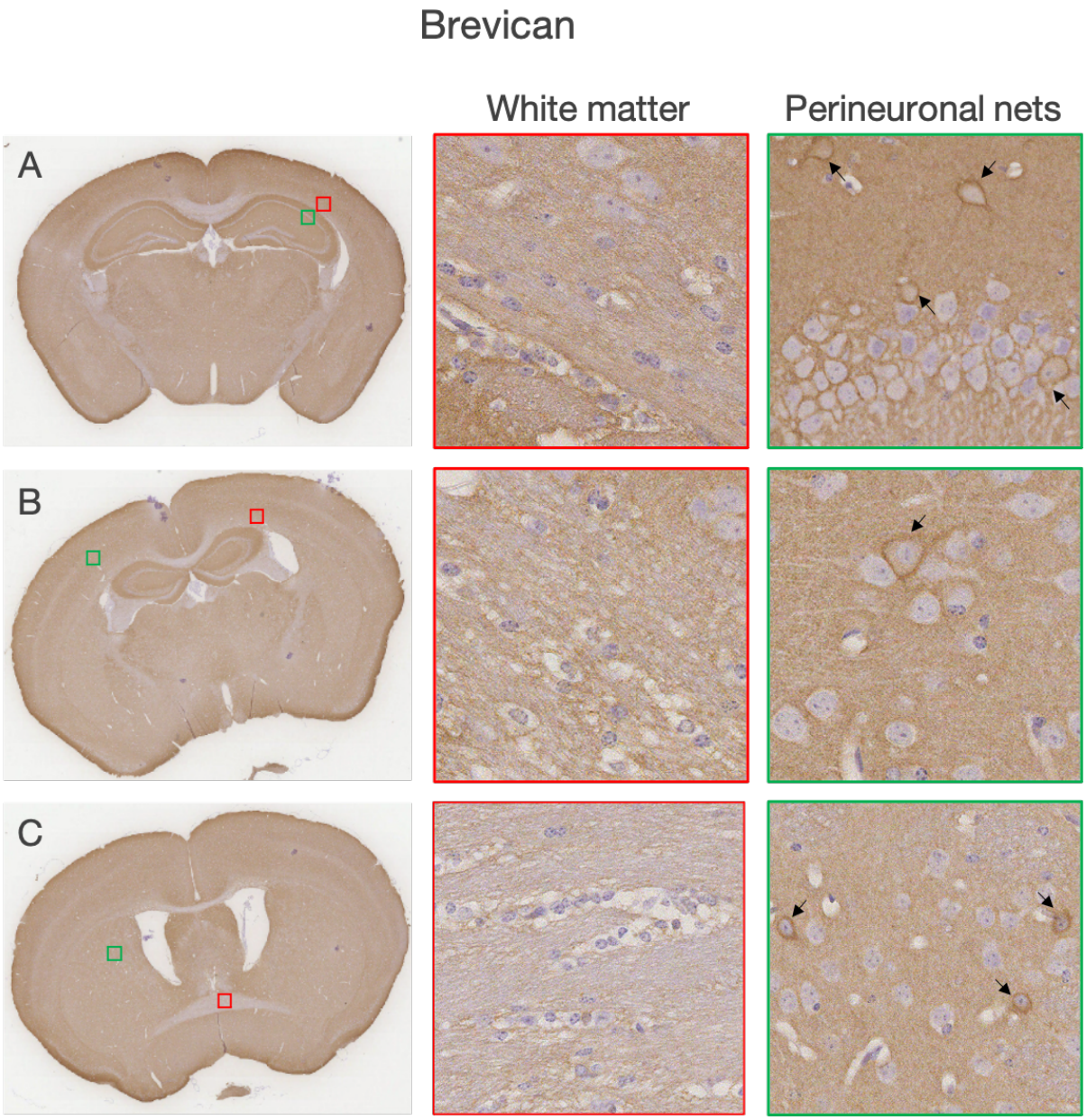

Figure 5.14. Example brevican stained brain slices (Abcam ab285162, 1:2000), digitised at the three locations of interest (A, B, C) in a wild-type mouse brain from the *Bcan* cohort, alongside zoomed-in inserts showing positive, diffuse staining in white matter (red) and staining accumulations in perineuronal nets (green, black arrows).

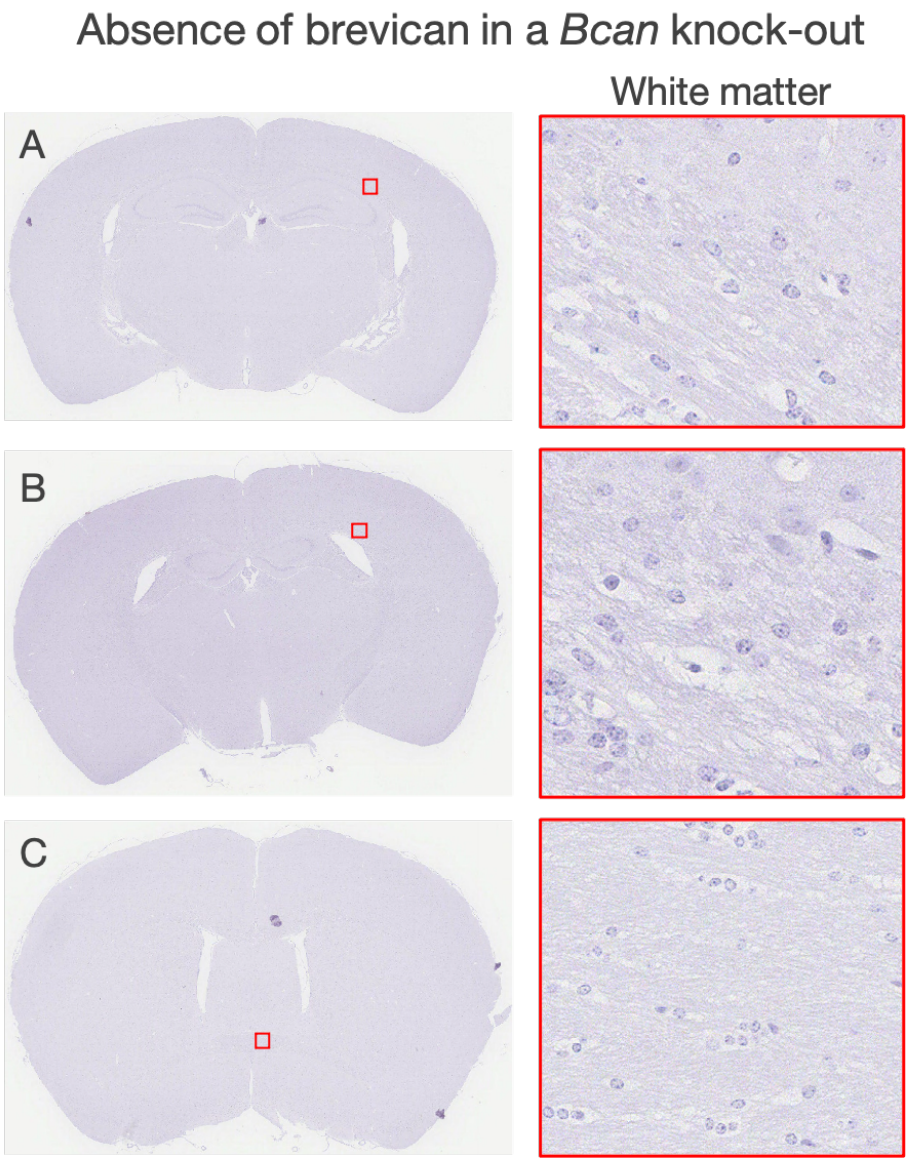

Figure 5.15. Absence of brevican staining in a *Bcan* knock-out mouse at the three locations of interest (A, B, C), alongside zoomed-in inserts of white matter. Brain slices stained using Abcam ab285162, 1:2000.

5.4. Quality control

We only selected brain slices presenting no major folds and tears for histological staining. Following staining, each slide was inspected under the microscope. Small artifacts such as DAB accumulation at the lower edge of the brain are to be expected. However, any artifact that would compromise SAF calculations in the regions which

will be included in the MRI-histology analysis deemed the slide unusable. In that case, the staining was then repeated using a neighbouring brain slice from the same location (A, B or C).

5.5. Histological data processing: methodology and results

5.5.1. Image Registration

We used FSL's TIRL (Huszar *et al.*, 2023) for the MRI-to-histology (v3.0) and histology-to-histology (v3.1) registration. The details of our multi-modal MRI datasets are described in **Chapter 4**. To account for the particularities of our datasets, minor adaptations were made to TIRL. This includes enabling it to accept new file formats (SVS files), evenly distributed control points, optimisation of search range extents and the ability to handle high-resolution MRI data. Our custom configuration files are available here: https://git.fmrib.ox.ac.uk/ihuszar/tirl/-/tree/v3/scripts/mouse/.

5.5.1.1. MRI-histology registration

The myelin PLP data was chosen as the reference histology data for MRI-histology registration due to its excellent grey-white matter contrast. This aids the registration algorithm, which relies on edges present in the MRI and histology data. For each individual mouse brain, the myelin PLP slides at each of the three locations of interest (A, B and C, **Figure 5.2**) were individually registered to the T_2-weighted MRI data acquired in the same brain. No photographed intermediaries are needed for the registration. The other stained slides (NF, brevican, versican V1, pan-versican) at each location of interest (A, B or C) for each individual mouse brain were registered directly to PLP.

Some pan-versican slides (8/24) were directly registered to MRI instead of myelin PLP slides, showing improved registration results. This was done because the pan-versican stain had to be repeated several times, resulting in stained slides which were further away from myelin PLP slides (40-50μm) than initially planned.

A configuration file was created for each registration. The manual input data required for the configuration file is listed in **Figure 5.16** c. The typical run time for each individual registration is around 30min.

The results of each registration run using the above default configuration were visually inspected (**Figure 5.16**, a), including evaluating the alignment of the three slices at the three locations (A, B, C, **Figure 5.16**, b). We expect the slices to be roughly parallel, although some deviations ($\pm$10°) are possible due to variations in the microtome adjustments during slicing of the histology block. In cases (6/48) where we observed a larger than expected deviation, we initialised a second registration using the outputs of the first registration but included a Gaussian smoothing step during the rigid-body registration stage. **Figure 5.16** a shows a typical MRI-histology registration result and the 3D visualisation of the registered histology slices for one sample at the three locations of interest (A, B and C).

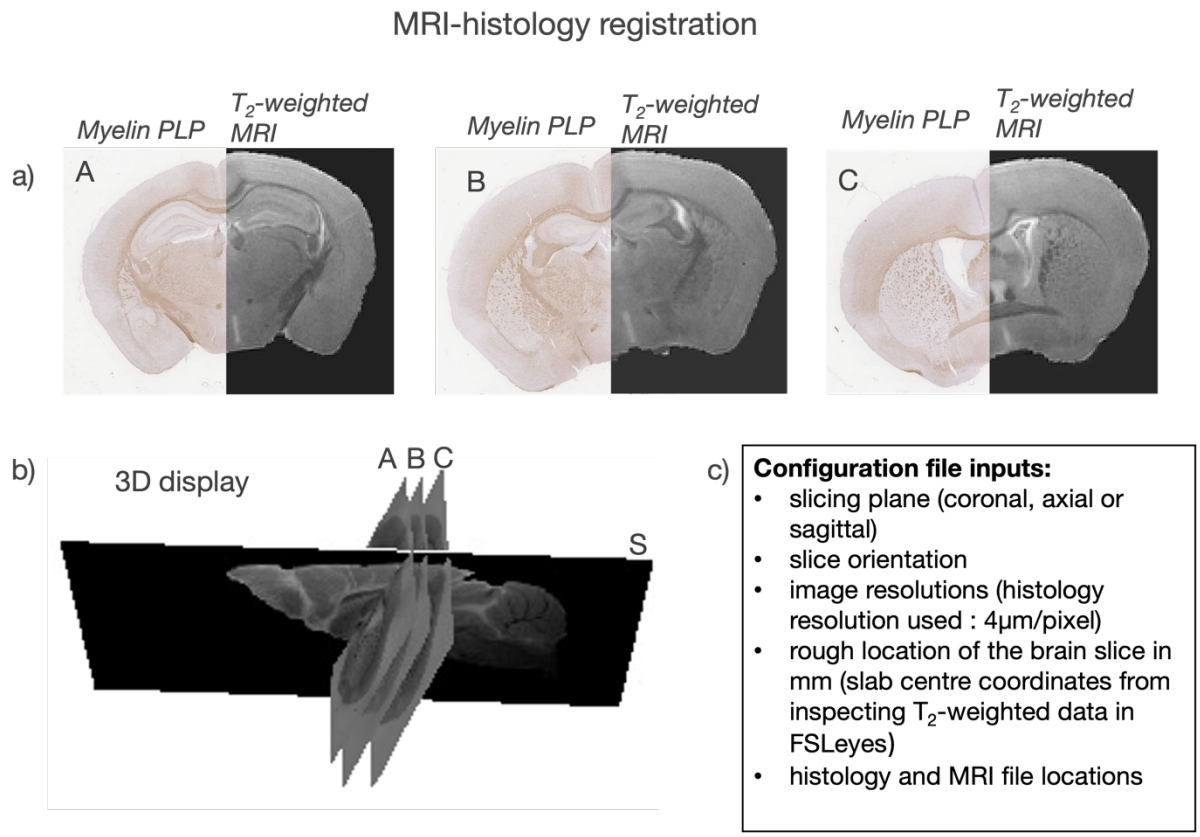

Figure 5.16. a) Example representative MRI-to-histology registration result in a wild-type mouse brain from the *Bcan* cohort, for the three locations of interest (A, B and C). b) 3D display of the three registered coronal brain slices (A, B, C). Sagittal plane (S): section through the 3D T_2-weighted structural MRI data.

We found that the final registration was highly accurate for all stained slices included in our study. The cost function reached an asymptotic minimum, and grey-white matter boundaries presented good alignment. While minor tissue tears and breaks did not appear to affect the MRI-histology registration, significant folds of the slice (**Figure 5.4**, a), lead to registration failure. Such data was not included in our study.

5.5.1.2. Histology-histology registration

A configuration file was created for each registration, with the manual inputs listed in **Figure 5.17**, b. The results of each individual registration were visually assessed. The registration quality was very accurate (**Figure 5.17**, a) and not affected by small tissue tears or folds. Histology-to-histology registration resulted in excellent alignment for all stains, including sparse ones (e.g., versican V1). This is because the histology-to-histology registration is guided primarily by the morphological similarity between the neighbouring slices and not by grey-white contrast.

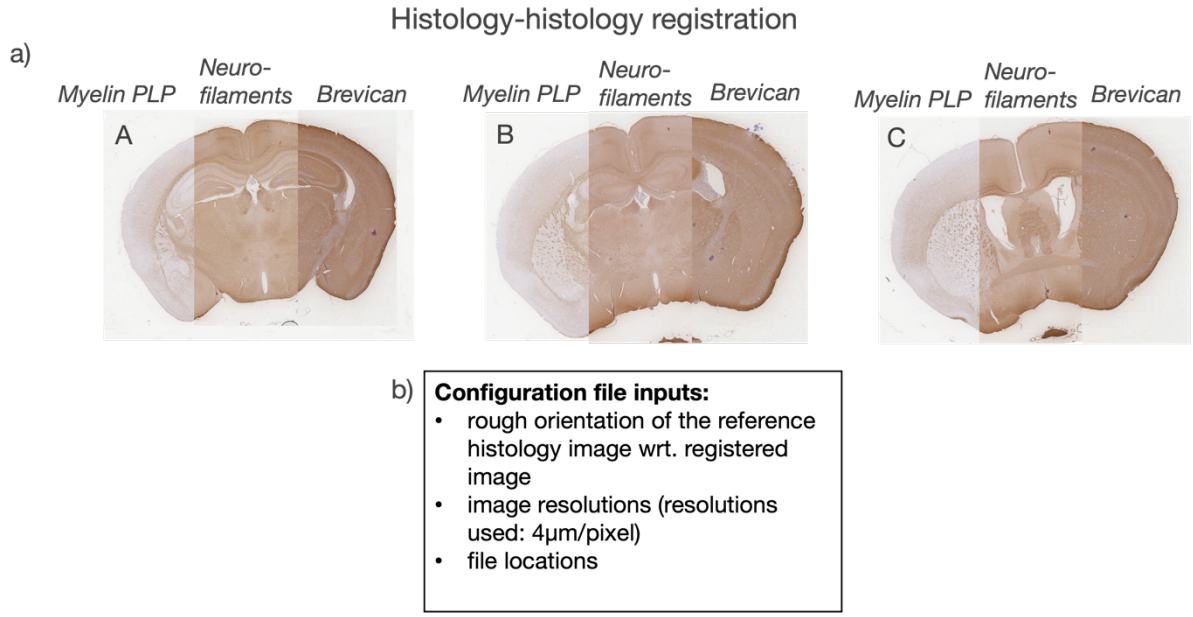

Figure 5.17. Example histology-to-histology registration output in a wild-type mouse brain from the *Bcan* cohort. The NF and brevican slides were separately registered to the myelin PLP slides at each of the three locations (A, B, C).

5.5.1.3. MRI-histology framework: registration chain

The MRI-to-histology and histology-to-histology registrations described above were embedded in a more comprehensive framework of co-registrations, including the other MRI modalities. Our multi-modal MRI to multi-stain histology registration chain is represented in **Figure 5.18**. Details of the FLIRT stages are presented in **Chapter 4**.

To obtain white matter masks of areas of interest, we used pydpiper (Chakravarty *et al.*, 2013; Friedel *et al.*, 2014) to perform iterative group-wise registration of the dMRI b=0 data to a cohort-specific common space, followed by atlas-based segmentation. The white matter masks were manually adjusted in standard space using the atlas-based segmentation output and the output from TBSS (Smith *et al.*, 2006) to ensure appropriate coverage of the white matter tracts of interest while avoiding partial volume effects. TIRL was used to concatenate the FLIRT transforms and resample the masks into the histology native space. This automatic approach overcomes the need for separate manual delineation of regions of interest on the histology data.

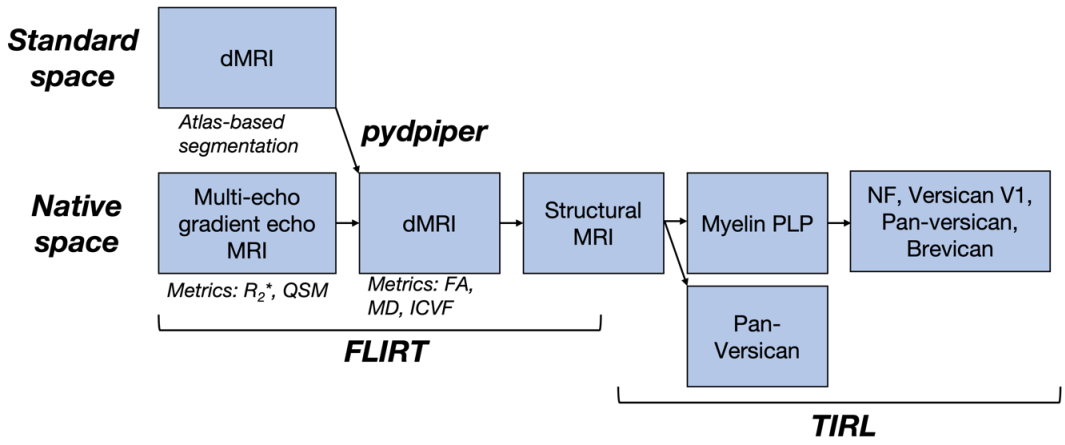

Figure 5.18. Multi-modal MRI to multi-stain histology registration framework. Pydpiper (Chakravarty *et al.*, 2013; Friedel *et al.*, 2014) is used for the cohort-specific common space registration and atlas-based segmentation of the dMRI data. FLIRT is used to co-register the multi-modal MRI data. TIRL is used for the MRI-to-histology and histology-to-histology registration.

5.5.2. Quantitative staining estimates

Following MRI-histology co-registration in each individual sample, we can directly correlate parametric estimates derived from multi-modal MRI (R_2^*, QSM, FA, MD, ICVF) to quantitative histological metrics at the MRI voxel level. Automatic derivation of the histology metrics is needed to ensure robustness and reproducibility of these results. Here, we use an automatic pipeline (Kor et al., 2022) to produce stain area fraction (SAF) maps for each digitised slide in our multi-stain histology dataset.

Stain Area Fraction pipeline

The first pipeline stage includes a stain separation step to isolate the DAB stain from the haematoxylin counterstain, which stains the cell nuclei purple. The common practice is to use a "fixed" colour matrix for all samples for each individual stain (Ruifrok and Johnston, 2001; Bagnato et al., 2018; Hametner et al., 2018). However, batch-to-batch variations may sometimes require an "adaptive" colour matrix estimated for each individual slide, which is implemented in the automatic SAF pipeline. The pipeline uses a k-means clustering approach (Geijs *et al.*, 2018) to calculate a 3x3 colour matrix that transforms the R, G and B colour channels to stain outputs (DAB, haematoxylin and residuals). We evaluated both options – "fixed" and "adaptive" – and found the best option to depend on the stain. Myelin PLP SAF maps showed particular improvement when using a fixed colour matrix (**Figure 5.19**). The other SAF maps were generated using the adaptive colour matrix.

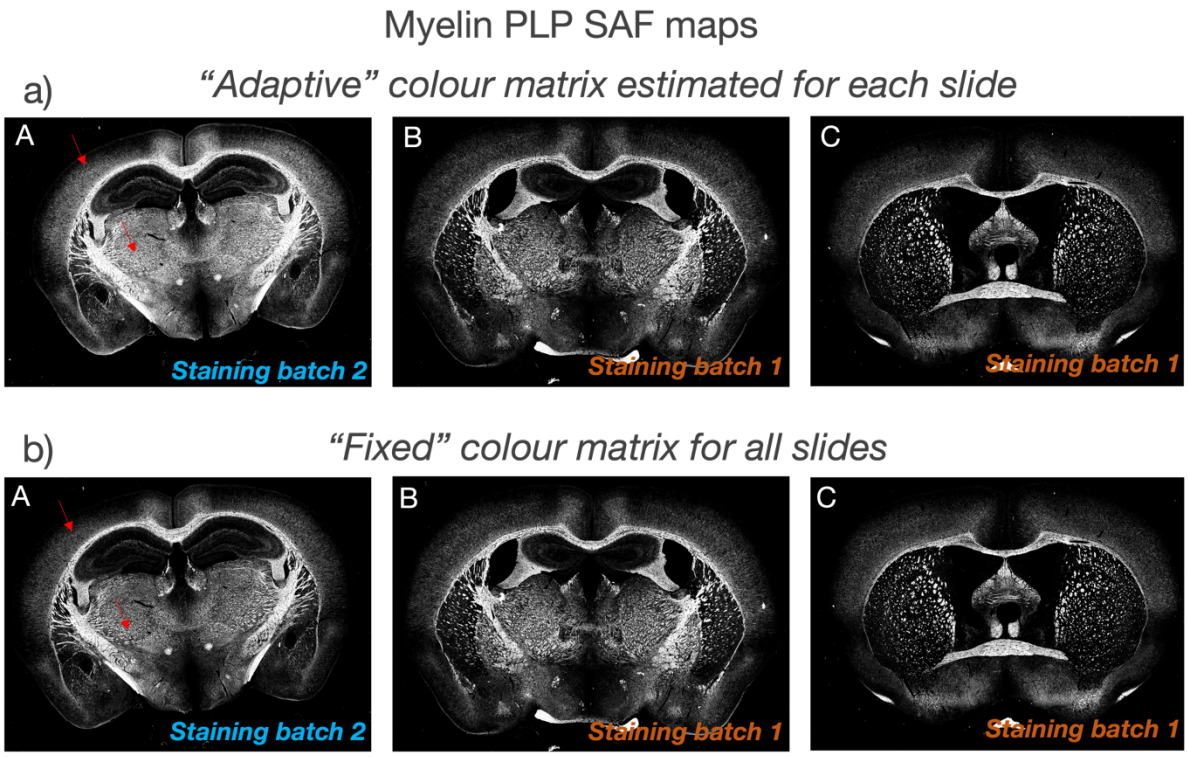

Figure 5.19. a) Example myelin PLP SAF maps obtained using an "adaptive" colour matrix per slide at the three slicing locations. Note that the SAF map obtained on the first section at location A appears brighter (red arrows), indicative of over-segmentation of the positive stain. b) Example myelin SAF maps obtained using a "fixed" colour matrix for all slides. Batch-to-batch variation in the SAF maps is reduced (red arrows).

Colour deconvolution is followed by the segmentation of the DAB-stained target protein from non-specific DAB staining, which requires manually tuning of a hyperparameter for each stain separately. The hyperparameter is determined empirically by evaluating the quality of the DAB segmentation across a range of values, avoiding over- or under-segmentation of the regions of interest (**Figure 5.20**). The hyperparameter allows for the computation of the threshold that maximises the image histogram's inter-class variance and is defined in the publication associated with the pipeline (Kor et al., 2022) . Finally, the SAF values are computed across each image patch. We used an 8 x 8 μm^2 patch size.

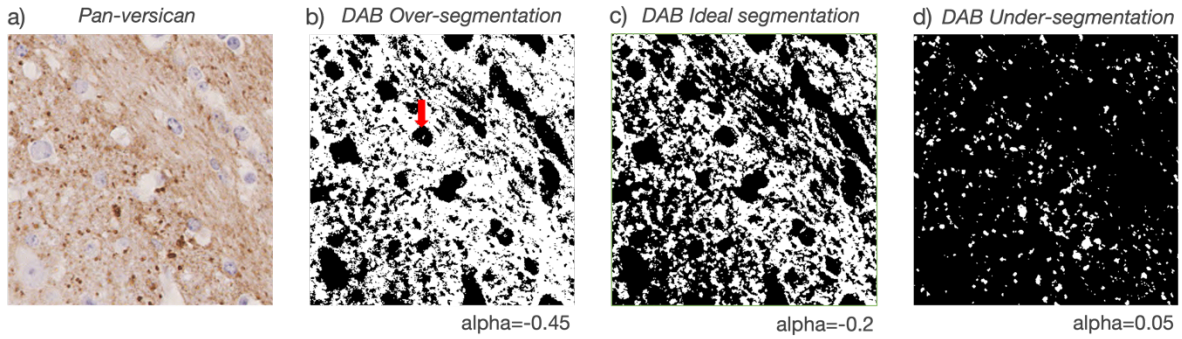

Figure 5.20. a) Patch showing pan-versican staining in white matter in a wild-type mouse brain from the *Vcan* cohort. b) Example DAB over-segmentation (hyperparameter alpha=-0.45), including false-positive segmentation inside cell bodies (red arrow) and background segmentation. c) Ideal DAB segmentation capturing the positive staining, including versican accumulation at the nodes of Ranvier and diffuse staining throughout white matter. Note that no staining is present inside cell bodies (denoted by the haematoxylin purple stain in a). d) Under-segmentation of the positive DAB stain. Only staining accumulations at the nodes of Ranvier are segmented, with positive staining in white matter not being included in the segmentation.

The hyperparameter values we chose are: 0.03 for NF, 0 for brevican, 0.05 for versican V1 and -0.20 for pan-versican. **Figure 5.21** shows representative SAF maps for all the stains. For the NF stains, we observed an artifact characterised by a strong edge effect on the inferior side of the brain slices, potentially due to the accumulation of DAB or antibodies at the edges of tissue. This cannot be corrected by the pipeline but it does not affect the regions of interest (white matter tracts) that will be included in the MRI-histology analysis. For the brevican stain, the SAF maps reveal the strong presence of brevican in the brain, surrounding neuronal cell bodies, with fainter, diffuse staining present in white matter. Finally, the pan-versican SAF maps present similar contrast to the myelin PLP SAF maps.

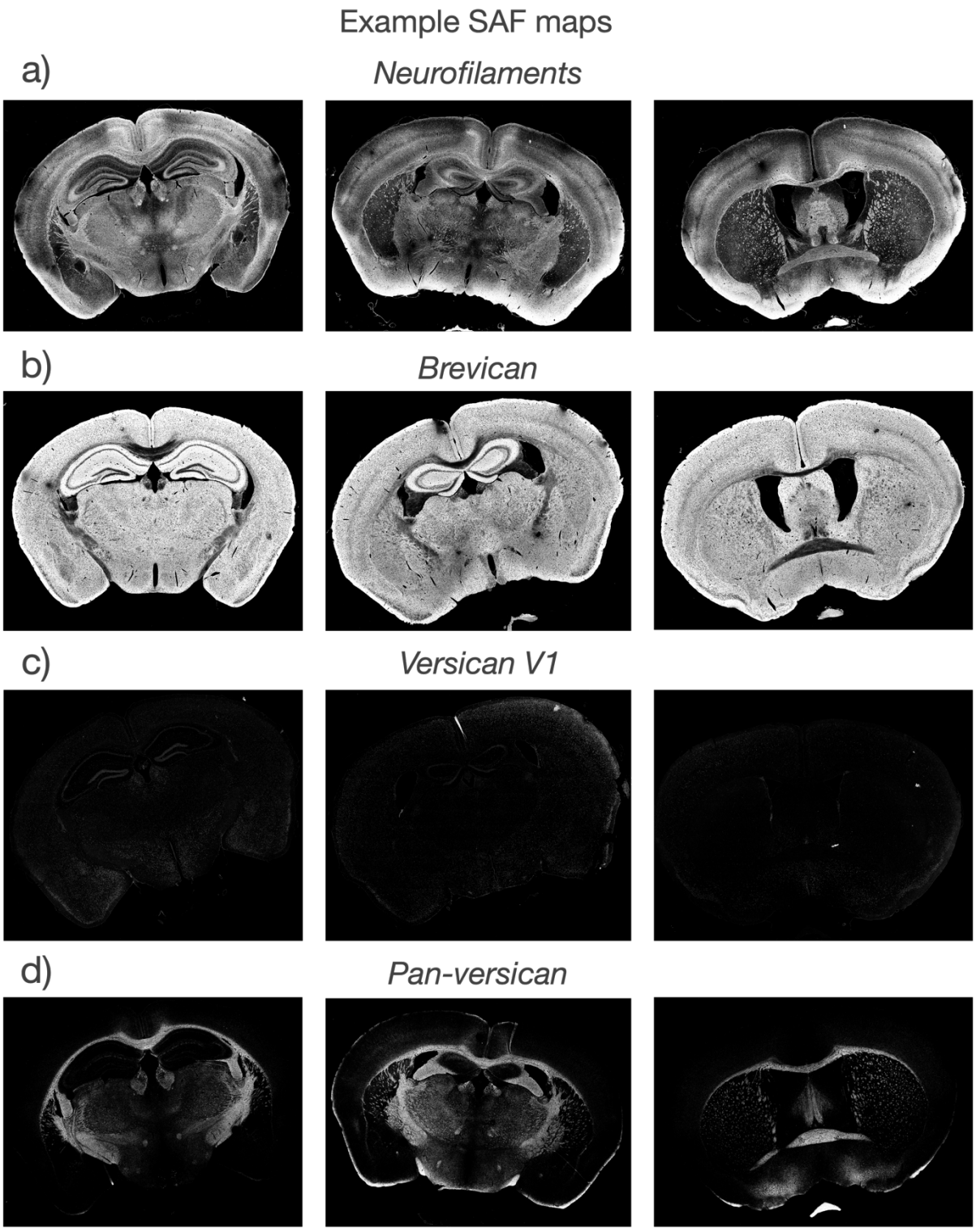

Figure 5.21. Example representative SAF maps at the three slicing locations. a) Example NF SAF map. b) Example brevican SAF map. c) Example versican V1 SAF map. d) Example pan-versican SAF map. All examples shown are from wild-type mouse brains from the *Bcan* cohort (NF, brevican) and *Vcan* cohort (versican V1, pan-versican).

5.6. Data open access

The histology data will be released on the Digital Brain Bank (Tendler *et al.*, 2022) and made available to external users following signing a material transfer agreement.

5.7. Discussion

To link multi-modal *ex-vivo* MRI with histology in the same mouse brain, we developed an optimised IHC protocol. This protocol accounts for the characteristics of mouse brain samples perfused for *ex-vivo* MRI, such as long storage times following perfusion fixation. This can both reduce antigenicity, affecting the quality of the stains, and render the samples brittle, which can result in severe morphological changes during histology processing, preventing registration to MRI. Our adapted protocol overcomes this by combining careful timing of the processing steps, choosing a slicing thickness that minimises folds and tears and devising IHC staining protocols that maximise the quality of each individual stain. Our resulting histology datasets present good stain quality and conserved brain morphology, allowing accurate MRI-histology data fusion.

The rodent imaging community is showing increasing interest in following up MRI studies with histology. Presently, histology protocols can vary across laboratories, their establishment and validation taking a significant amount of time. To address these reproducibility challenges, we aimed to make our protocols and processing strategies available to the wider community. The step-by-step details of our protocols are highlighted in our publication (Smart *et al.*, 2023). While the protocols may not be suitable for freshly prepared tissue or larger specimens, we expect them to yield comparable results for similarly-sized tissue samples prepared for *ex-vivo* MRI.

We found that artifacts can arise at all histological processing stages and recommend strict adherence to optimised timings. It is also important to account for the random presence of artifacts. Random staining artifacts, such as strong edge effects due to DAB/antibody accumulations or fainter staining due to air bubbles, which affect major

sections of the brain slice, occur on approximately 25% of all slides. In such cases, neighbouring slides will have to be used for re-staining.

Tissue shrinkage and expansion occurs at all stages of processing, through the paraffin-embedding (alcohol to xylene to paraffin) and IHC staining (paraffin to xylene to ethanol to rehydration) processes. We expect the stained brain slices to be of roughly similar size to the MR-imaged brain volumes, with the slice morphology broadly preserved. The registration tool we use (Huszar *et al.*, 2023) will account for small non-linear tissue deformation, as well as 3D slice deformations. This tool has been subjected to a wide range of tests to assess its accuracy (Huszar *et al.*, 2023). We found that it only fails when severe morphological alterations to the brain slice are present, such as folds across the cortical surface. This is the primary issue to be avoided during histological processing. Folds and large tears can be observed during tissue slicing and solutions should be devised at this stage to mitigate them, some of which we highlight in our paper (Smart *et al.*, 2023).

5.8. Conclusions

This chapter presented our work on developing an FFPE-IHC protocol for mouse brains prepared for *ex-vivo* MRI, which maintains the morphology of brain slices and yields high quality myelin PLP, NF, brevican, versican V1 and pan-versican stained brain slices. This data is optimal for MRI-histology and histology-histology registration and for the extraction of SAF quantitative maps.

Chapter 6: Phenotyping of the *Vcan* and *Bcan* mouse models using multi-modal *ex-vivo* MRI and histology

Context

In **Chapter 3**, we introduced the neuroimaging component of UK Biobank and the key results: associations between multiple imaging-derived phenotypes (IDPs) and single nucleotide polymorphisms (SNPs) in the *VCAN* and *BCAN* genes, which encode extracellular matrix (ECM) proteins. This chapter delves into the UK Biobank results, proposing an approach to explore the causal relationships between variations in the *VCAN* and *BCAN* genes and the differences in IDPs. To achieve this, we phenotype mouse models of *Vcan* and *Bcan* using the MRI and histology protocols described in **Chapters 4** and **5**. Here, we describe the preliminary findings yielded by these datasets.

Published work

Personal contributions

Conference Proceedings

Cristiana Tisca, Mohamed Tachrount, Frederik J Lange, Amy FD Howard, Chaoyue Wang, Lily R Qiu, Benjamin C Tendler, Jason P Lerch, Aurea B Martins-Bach, Karla L Miller. *White matter microstructure changes in a Bcan knock-out mouse model.* ISMRM & ESMRMB Joint meeting (2022).

Cristiana Tisca, Mohamed Tachrount, Chaoyue Wang, Lily Qiu, Javier Barallobre-Barreiro, Marika Fava, Manuel Mayr, Jason Lerch, Aurea B Martins-Bach, Karla L Miller. *Vcan mutation leads to sex-specific changes in white matter microstructure in mice*. ISMRM & SMRT Virtual Conference (2021).

Individual contributions to this Thesis Chapter

My personal contributions to this chapter include acquiring, processing, evaluating, and analysing the multi-modal MRI and multi-stain histology for the *Vcan* and *Bcan* mouse cohorts.

The *Vcan* mouse model was developed by MRC Harwell and King's College London and kindly provided by Dr Javier Barallobre-Barreiro. The *Bcan* mouse model was developed by MRC Harwell as part of the International Mouse Phenotyping Consortium. All samples were perfusion-fixed at MRC Harwell.

6.1. Introduction

The MR neuroimaging component of UK Biobank acquires brain imaging data in very large populations. The first imaging genome-wide association studies (GWAS) exploring this UK Biobank data (Elliott et al., 2018) used multi-modal MRI data from ~10,000 participants to link ~3000 IDPs to ~11 million SNPs. Many different genetic regions associating with IDPs were identified. More recently, data from 40,000 participants was analysed (Smith *et al.*, 2021), with results made publicly available (https://open.win.ox.ac.uk/ukbiobank/big40/). Among these studies, some of the strongest and most consistent results relate a range of IDPs to SNPs in the *VCAN* and *BCAN* genes. A separate GWAS, investigating QSM and R_2^* phenotypes, found associations between susceptibility (χ) in white matter hyperintensities and a SNP in *VCAN* (Wang *et al.*, 2022). A summary of the associations between SNPs in *VCAN* and *BCAN* and IDPs revealed in these two studies is presented in **Table 6.1**.

Modality	IDP category	*VCAN* (total hits: 578)		*BCAN* (total hits: 31)	
		#hits	top p	#hits	top p
dMRI	ICVF	140	1.69e-76	23	4.74e-13
	ISOVF	13	2.42e-17		
	OD	10	3.71e-12		
	FA	65	3.58e-44	4	4.49e-10
	MD	80	2e-44		
T_1-weighted	Gray/white matter contrast	34	8.72e-29		
QSM	χ in WMH	2	1.77e-19		

Table 6.1. Top UK Biobank GWAS summary results (discovery dataset) using the 40,000-subject dataset. ICVF: intra-cellular volume fraction; FA: fractional anisotropy; MD: mean diffusivity; QSM: quantitative susceptibility mapping; χ: susceptibility; WMH: white matter hyperintensities; top p: highest p-value among hits in the IDP category. Total hits: total number of statistically-significant associations identified. This includes but is not limited to the listed top hits. Note: Blank cells indicate no hits for a given category of IDPs. The data was pooled from https://open.win.ox.ac.uk/ukbiobank/big40/.

VCAN and *BCAN* encode versican and brevican, respectively, which are ECM proteins belonging to the same class with similar roles in the brain. The various roles that versican (through its isoforms) and brevican have in the development and function of the brain are presented in **Chapter 2**. In particular, versican and brevican are enhanced during white matter development and following injury and they inhibit myelin regeneration (Lau *et al.*, 2013) and neurite outgrowth (Beller and Snow, 2014).

Despite these known roles of versican and brevican in the brain, the biological mechanisms leading to the specific signal changes in MRI reported in **Table 6.1** remain unclear. In particular, the ECM is not expected to be a direct driver of the specific MRI signals captured by the GWAS hits (ICVF, FA, MD, T_1 grey-white contrast or χ).

In **Chapter 2**, we introduced the approach of using mouse models to further explore results from GWAS research. We propose to use this reverse translational approach to investigate the causal relationships between variations in the *VCAN* and *BCAN* genes and changes in IDPs. To do this, we use *ex-vivo* multi-modal MRI and multi-stain immunohistochemistry (IHC) to characterise the brains of two *Vcan* and *Bcan* mouse models using the protocols described in **Chapters 4** and **5**.

We do not expect to observe the exact same phenotypes as in UK Biobank. We cannot engineer a mouse model where we replicate the exact effects of a human SNP, due to cross-species genetic and physiological differences. Researchers frequently target the specific gene more broadly, either by fully knocking it out, or via more sophisticated genetic engineering techniques.

Nevertheless, one important benefit of working with mouse models is that cohorts can be extremely homogeneous. In our studies, mice belonging to the same cohort have the same genetic background (C56BL/6), age, parents and environment. This lower phenotypic variance in mice bridges the gap between the large cohort sizes in humans and much smaller cohorts in mice.

For these reasons, we hypothesise that our *Vcan* and *Bcan* mouse models will display similar and not identical phenotypes to those observed in UK Biobank. We expect that a genetically-driven microstructural alteration in white matter can drive changes in one or more of the metrics computed via dMRI (DTI and NODDI), QSM and R_2^*.

Mouse models of *Vcan* and *Bcan* have not been previously studied using multi-modal MRI. Moreover, while some histology studies do exist of these models, they only provide a qualitative description of the stains (Brakebusch et al., 2002; Dours-Zimmermann et al., 2009). Employing a brain-wide, quantitative characterisation of these models using MRI and histology can offer new insights into how changes in the versican and brevican proteins affect MRI phenotypes.

In this chapter, we acquire and analyse MRI and histology data from the *Vcan* and *Bcan* mouse models. We first quantify how multiple microstructural MRI metrics are differentially affected in the *Vcan* and *Bcan* mouse cohorts. Differences are observed in both cohorts, particularly in R_2^* and grey-white matter contrast in male *Bcan* mice. These changes relate to those observed in UK Biobank.

To pinpoint the specific biological cause of these changes, we acquire multi-stain IHC histology in a subset of brains from these cohorts and process this data to output a quantitative histology metric (SAF). We hypothesise that the observed changes in the MRI signal are due to genetically-driven differences in myelin proteolipid protein (PLP), neurofilaments (NF) or both. We conduct qualitative and quantitative analyses of the IHC data in a subset of mice. While we do not observe any major changes in histology in either the *Vcan* or *Bcan* cohorts, some consistent trends are starting to emerge.

6.2. Materials and methods

In sub-sections 6.2.1. and 6.2.2., we will first describe the UK Biobank GWAS results in more detail, highlighting the limitations that we have in understanding the causal mechanisms driving the observed changes in IDPs in humans. We will then introduce the genetic modifications present in our mouse models, followed by a description of key findings in previous studies of *Vcan* and *Bcan* mouse models. We subsequently describe the data acquisition and processing and the statistical analyses we conducted.

6.2.1. The *Vcan* mouse model

The strongest genetic effect identified in the first imaging GWAS (Elliott et al., 2018) was between a SNP in an intron – non-coding region – of *VCAN* and 199 brain-wide white matter dMRI phenotypes (NODDI's ICVF and DTI's FA and MD). This SNP is in a "distal enhancer-like signature" region, according to UCSC's genome browser (https://genome.ucsc.edu/). This type of region possibly regulates transcription (Stadhouders *et al.*, 2012), potentially leading to differential expression of versican. It is unclear which versican isoform is affected by this process. It is also unclear whether the UK Biobank hits associate are due to enhanced or reduced expression of versican, or more complex biological mechanisms.

Knocking out the *Vcan* gene in mouse models leads to embryonic lethality (Mjaatvedt *et al.*, 1998). Mouse models with modifications in *Vcan* thus require more precise genetic modifications. Our *Vcan* mouse model (background: C56BL/6) has a point mutation (E441A). This mutation is in a location where the versican V1 isoform gets broken down – or cleaved – and it leads to impairment in this process. Due to this genetic modification, there is a possible accumulation of the versican V1 isoform. This could possibly interact with the other versican isoforms. No phenotypic characterisation of this model has been released yet.

Other *Vcan* mouse models with different genetic modifications have been previously described in literature. Researchers have knocked-out only the expression of V0/V2 isoforms (Dours-Zimmermann et al., 2009). These mice were viable and fertile. Their brain and white matter development was not affected by the lack of versican V2. The distribution of myelin basic protein was not visibly affected; however, the ECM at the nodes of Ranvier was severely altered.

A recent study described a *Vcan* model with two point mutations (E441A – the one our mouse model has – and E438A) (Nandadasa *et al.*, 2021). This model did not show severe phenotypes but showed changes in limb development. No differences in brain structure were reported.

6.2.2. The *Bcan* mouse model

In the first UK Biobank neuroimaging GWAS (Elliott et al., 2018), a SNP in an eQTL of *BCAN* – a region regulating genetic expression – was associated with ICVF in the genu of the corpus callosum. Likewise, it is unclear whether these UK Biobank hits are due to an enhancement or reduction in brevican expression, or other mechanisms.

The *Bcan* mouse model (background: C56BL/6) is a knock-out model, lacking brevican. The genetic modification introduces a "premature stop codon" (spanning the first two exons of the 201 and 205 transcripts), which prevents the brevican protein from being translated. This mouse model was developed by MRC Harwell within the IMPC study (https://www.mousephenotype.org/data/genes/MGI:1096385). Some phenotypic tests have been conducted within IMPC, but no severe phenotypes were observed. In phenotyping tests done at MRC Harwell, *Bcan* knock-outs showed hypoactivity in the open field test in both males and females and abnormal maxilla in males only. The MRC Harwell phenotyping results have not yet been officially released or statistically analysed.

Bcan knock-outs have been characterised in literature, showing no obvious changes in brain morphology (Brakebusch et al., 2002). However, they show impaired synapse function in the hippocampus (Brakebusch et al., 2002). *Bcan* knock-out mice also show more extensive neurophysiological alterations in neurons with perineuronal nets and associated impairments in working and short-term memory (Favuzzi *et al.*, 2017).

6.2.3. Sample preparation, data acquisition and processing

The *Vcan* and *Bcan* mice were perfusion-fixed (described in **Chapter 4**) at 10-11 weeks and the brain samples were kept in the storage solution at 4°C until scanning. This age was chosen to reduce variance due to developmental factors as mice enter adulthood around that time. The samples were scanned using the multi-modal MRI protocols presented in **Chapter 4**. Due to COVID-19-related delays, scanning was done 9-17 months after perfusion fixation.

There was no pilot data available for the *Vcan* or *Bcan* cohorts. Therefore, the expected effect size could not be estimated, and a power analysis calculation was not conducted. The cohort size was therefore estimated based on other similar mouse studies from our centre. The cohorts consist of four groups: male homozygous (M HOM), male wild-type (M WT), female homozygous (F HOM) and female wild-type (F WT). Each group was allocated 7 samples, totalling 28 mice/cohort. Due to sample damage partway through the MRI data acquisition, the available sample size was lower than initially allocated. Cohort distributions are presented in **Table 6.2**.

2 samples per group for each of the *Vcan* and *Bcan* cohorts were selected for multi-stain IHC histology using the protocols described in **Chapter 5**. This size was limited to the time available during the thesis research and is expected to be increased in post-doctoral work. Histology was performed ~27 months after perfusion-fixation. Both cohorts were stained for myelin PLP and NF. The mice from the *Vcan* cohort were also stained for versican V1 and pan-versican (essentially a versican V2 stain, see

Chapter 5). The mice from the *Bcan* cohort were also stained for brevican. Examples of stained slides for each individual stain are shown in **Chapter 5**.

Cohort	Data	# M-HOM	# M-WT	# F-HOM	# F-WT
Vcan	T_2-weighted	7	7	7	7
	dMRI	7	7	7	7
	Multi-echo GRE	6	5	5	6
	IHC Histology	2	2	2	2
Bcan	T_2-weighted	7	7	7	5
	dMRI	7	7	7	5
	Multi-echo GRE	7	7	7	5
	IHC Histology	2	2	2	2

Table 6.2. Cohort distributions for the *Vcan* and *Bcan* cohorts. Due to sample damage partway through the MRI acquisition, the initial cohort size of 7:7:7:7 (in green) was not available for all imaging experiments. Only 22 *Vcan* mice were available for multi-echo GRE data acquisition (in orange). 26 Bcan mice were available for MRI data acquisition (in yellow). We used a subset of 2 mice/group for histology processing.

The dMRI and multi-echo GRE data were processed using the pipelines described in **Chapter 4** to extract FA, MD, ICVF, R_2^* and QSM maps. If acquired in separate sessions, the MRI data was co-registered (as described in **Chapter 4**). The histology data was processed to extract stain area fraction (SAF) maps (as described in **Chapter 5**) and registered to MRI (as described in **Chapter 5**). Neighbouring histological slides were co-registered (as described in **Chapter 5**).

The T_2-weighted structural data is used to extract the grey-white matter ratio across an eroded isocortex mask (as detailed in **Chapter 4**).

The T_2-weighted structural data and the averaged dMRI b=0 data for each cohort were registered to a cohort-specific standard space and segmented based on the DSURQE

mouse brain atlas (Chakravarty *et al.*, 2013; Friedel *et al.*, 2014). White matter tract masks were manually adjusted in the dMRI cohort-specific standard space using the atlas segmentation and a TBSS skeleton (Smith *et al.*, 2006). This ensured appropriate coverage of white matter tracts, avoiding partial volume effects. The tracts included in the MRI and histology analyses are displayed in standard and native space, respectively (**Figure 6.1**). In QSM, myelin anisotropy induces susceptibility variations depending on the orientation of the white matter fibres with respect to B_0 (Xu *et al.*, 2018). The white matter masks used for the QSM analysis therefore only contained voxels where tracts lay within the 30° of the plane perpendicular to the B_0, calculated from the principal diffusion direction. Only the anterior commissure, corpus callosum and fimbria had a significant proportion of the tract satisfying that condition, so only those tracts were included in the QSM analysis.

The white matter segmentations were mapped through to histology native space. The fornix and cerebral peduncle were excluded from the histology analysis due to their absence in individual slices in some of the samples. Ventricle expansion during histological processing led to some tract shrinkage. Hence, these tracts are over-segmented in the histology native space (**Figure 6.1** b, location B), white arrow). To avoid including non-white matter regions from histology, we only incorporated voxels which showed non-zero SAF values in the analysis.

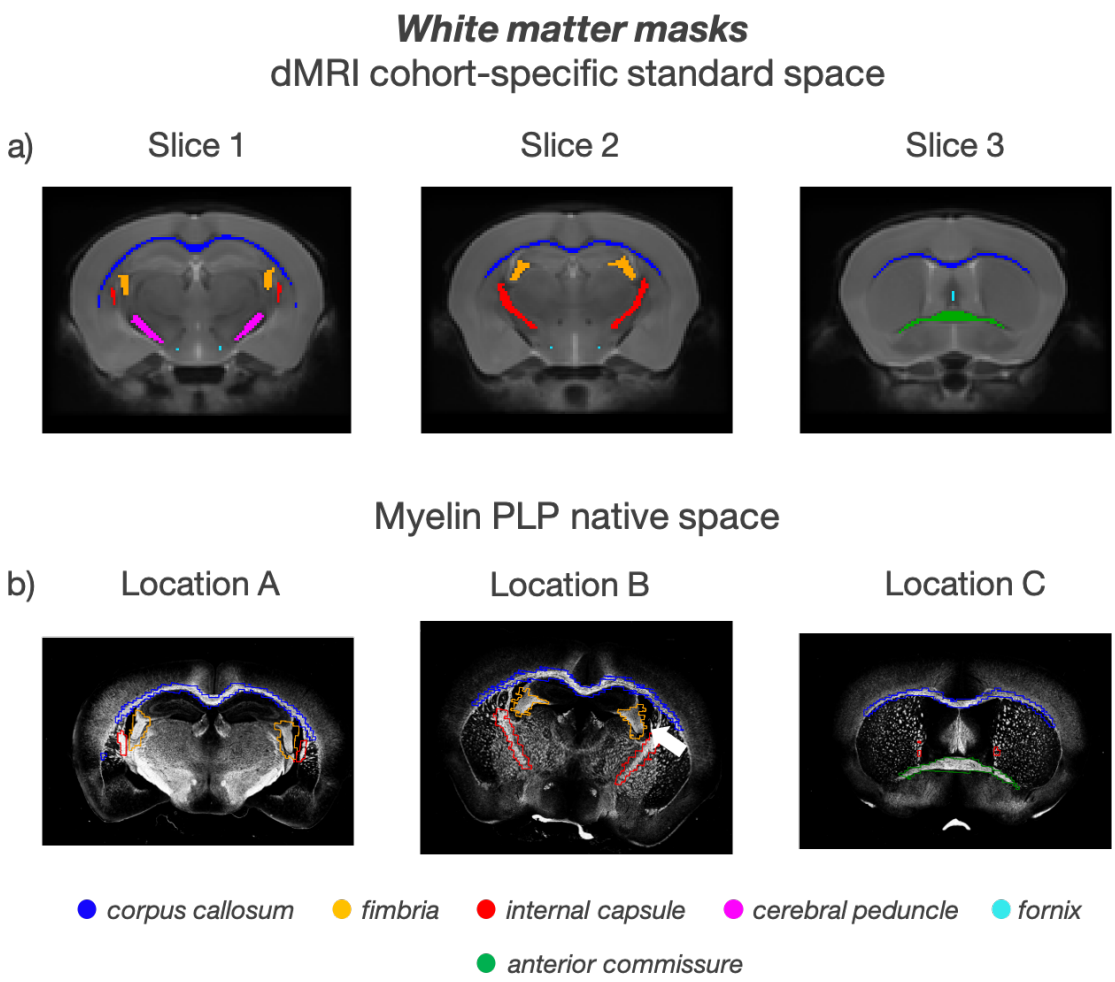

Figure 6.1. White matter masks in a) dMRI cohort-specific standard space and b) myelin PLP native space. Slices 1-3 in myelin PLP native space correspond to locations A-C at which we obtained slices in each mouse brain (described in **Chapter 5**). The white arrow in "Slice 2" shows over-segmentation of the fimbria, occurring due to tract shrinkage compared to MRI.

6.2.4. Statistical analysis

6.2.4.1. MRI

We used a linear mixed-effects model to estimate the effect of genotype and sex on brain volume (either voxel-wise or for each segmented region in the atlas) using the

absolute Jacobian determinant, which measures local expansions and contractions as a function of genotype. Multiple comparison correction was done using false discovery rate correction (Genovese, Lazar and Nichols, 2002), thresholded at 15%.

The microstructural MRI analysis on FA, MD, ICVF, R_2^* and QSM was conducted in each modality's native space. We computed the median value across each individual tract shown in **Figure 6.1** a. The median value of the grey-white matter ratio within the isocortex mask was calculated for each individual sample as well and incorporated into the statistical analysis. Median values were compared between HOMs and WTs, accounting for sex. FSL's Permutation Analysis of Linear Models (PALM) (Winkler *et al.*, 2014) was used for non-parametric permutation testing (5000 iterations) across the 4 contrasts, 6 metrics (FA, MD, ICVF, QSM, R_2^* and grey/white matter ratio) and 6 tracts of interest. PALM facilitates inferences across multiple metrics, using family-wise error rate (FWER) to correct for multiple comparisons.

6.2.4.2. MRI-histology analysis

6.2.4.2.1. Simple correlation

We correlated each pair of SAF metrics (PLP, NF, versican and brevican) to MRI metrics (FA, MD, ICVF, R_2^* and QSM) across the white matter tracts of interest using Pearson correlation, across WTs and HOMs from both cohorts.

6.2.4.2.2. Composite modelling

We developed a composite model which allows us to separate the relative contributions to differences in MRI that are due to genetics (**Figure 6.2**). We hypothesise that the observed changes in each individual MRI metric can be explained either directly by genotype or via differences in NF or PLP SAF, which include genetically- (g) and non-genetically (ng)-driven components. This composite model

allows us to calculate how much the genetic alterations in MRI metrics are due to genetically-driven changes in NF and myelin PLP.

Here, $\beta_{g,PLP}$ and $\beta_{g.NF}$ are the coefficients estimated using regression, g is the genotype (HOM and WT). PLP_g and NF_g are the genetic components of the myelin PLP and NF stains, respectively. PLP_{ng} and NF_{ng} are the model residuals. To enable fitting of the composite model, all three non-genetic components (PLP_{ng}, NF_{ng} and MRI_metric_{ng}) cannot correlate with each other. Given the likelihood that sex and tract, in addition to genotype, lead to variations in PLP, NF and the individual MRI metrics, we cannot pool the data across sexes and tracts at this stage.

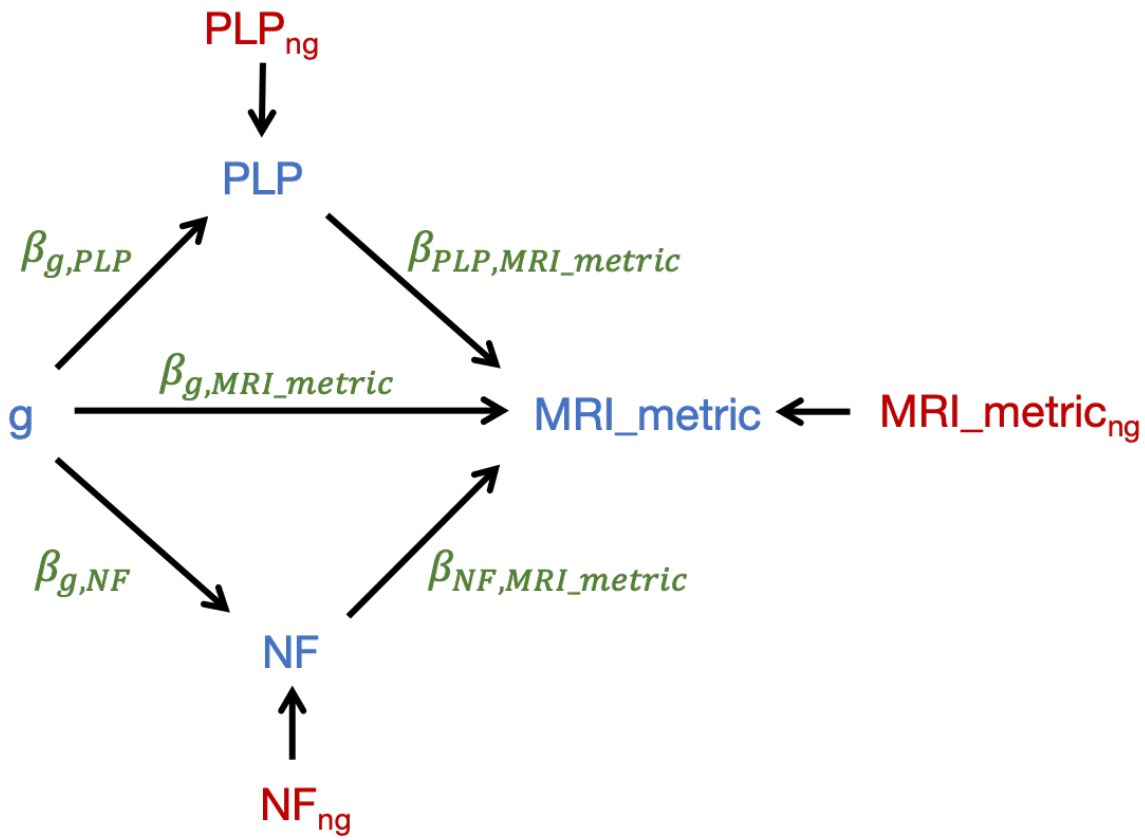

Figure 6.2. Modelling the genetic contributions to each MRI metric either via myelin PLP or NF. g: genotype; ng: non-genetic contribution.

Equations 6.1-6.2 describe the model in **Figure 6.2**. We fit each equation separately to estimate the five β coefficients, and then combine the equations to gain insight into the different pathways by which genetics might be influencing the individual MRI metrics.

$$PLP = PLP_g + PLP_{ng} = \beta_{g,PLP} g + PLP_{ng}$$

(Equation 6.1.)

$$NF = NF_g + NF_{ng} = \beta_{g,NF} g + NF_{ng}$$

(Equation 6.2.)

$$MRI_metric = \beta_{g,MRI_metric} g + \beta_{PLP,MRI_metric} PLP + \beta_{NF,MRI_metric} NF$$

(Equation 6.3.)

We conducted a preliminary analysis employing this model using voxel-wise MRI-histology data from the fimbria (location B) in *Bcan* males, where we observed the strongest effects across modalities. The model was fitted separately for each MRI metric. All data is demeaned prior to fitting.

Using this composite model, we then calculated the fraction of a given MRI metric that is explained by genetics, e.g., $std(\beta_{g,MRI_metric} + \beta_{PLP,MRI_metric}\beta_{g,PLP} + \beta_{NF,MRI_metric})\beta_{g,NF})/std(MRI_metric)$. In addition, we calculated the genetic contributions to PLP ($(std(\beta_{g,PLP} g)/std(PLP)$) and NF ($std(\beta_{g,NF} g)/std(NF)$).

6.3. Results

6.3.1. MRI

6.3.1.1. Volumetric analysis

No significant differences in brain volume (at the voxel or ROI level) between HOMs and WTs were detected in the volumetric analysis of either the *Vcan* or *Bcan* cohort.

6.3.1.2. Grey-white matter ratio

Significant differences in median grey-white matter ratio (p-value=0.005) were observed between males in the *Bcan* cohort (**Figure 6.3** b), with HOM males showing lower values than WTs. An opposite, although not significant (p=0.092) effect was observed in *Bcan* females, with higher values in HOM mice compared to WTs.

No statistical differences in this metric were detected in the *Vcan* cohort (**Figure 6.3** a). However, an inverse trend compared to *Bcan* was visible, with higher values in *Vcan* male HOMs compared to WTs, and lower in *Vcan* female HOMs compared to WTs.

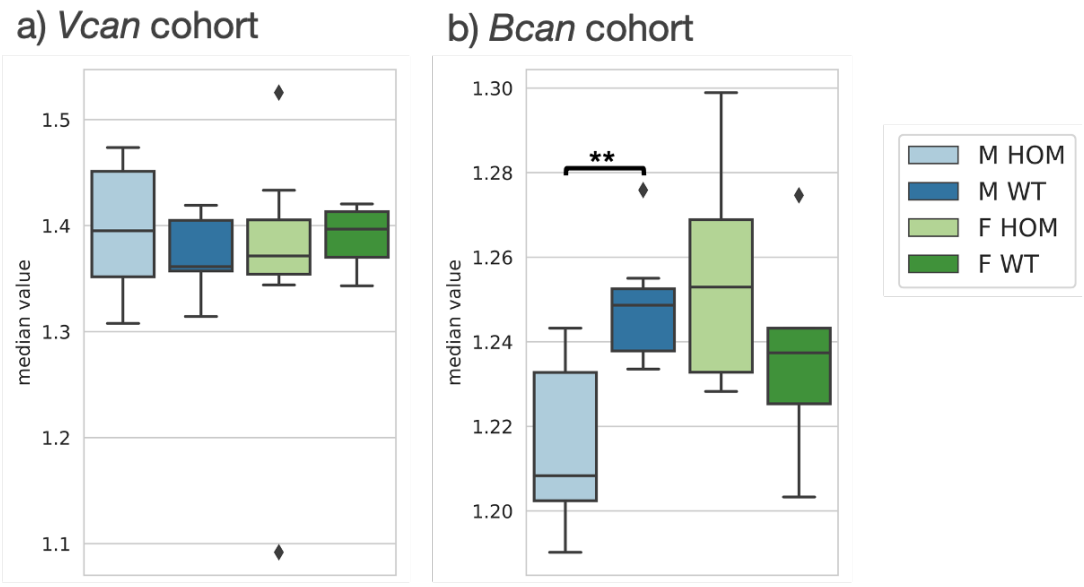

Figure 6.3. Median grey-white matter ratio values across an isocortex mask, as extracted from the T_2-weighted structural data. M HOM: male homozygous; M WT: male wild-type; F HOM: female homozygous; F WT: female wild-type. Boxplot components: edges – inter-quartile range, middle line – median value, whiskers – data range (reaching to $\pm$ 1.5 beyond the first and third quartiles, respectively), flier points – outliers past the end of the whiskers. Plots produced with Matplotlib v3.3.4.

6.3.1.3. Diffusion-weighted MRI

Boxplots of median values for FA, MD and ICVF, for each individual group (M HOM, M WT, F HOM and F WT) and for both cohorts are presented in **Figure 6.4**. Consistently lower FA, MD and higher ICVF are detected in *Bcan* HOM males compared to WTs, although the differences are not statistically significant. Opposite trends are observed in the *Vcan* cohort, with consistently higher FA values in male HOMs compared to WTs.

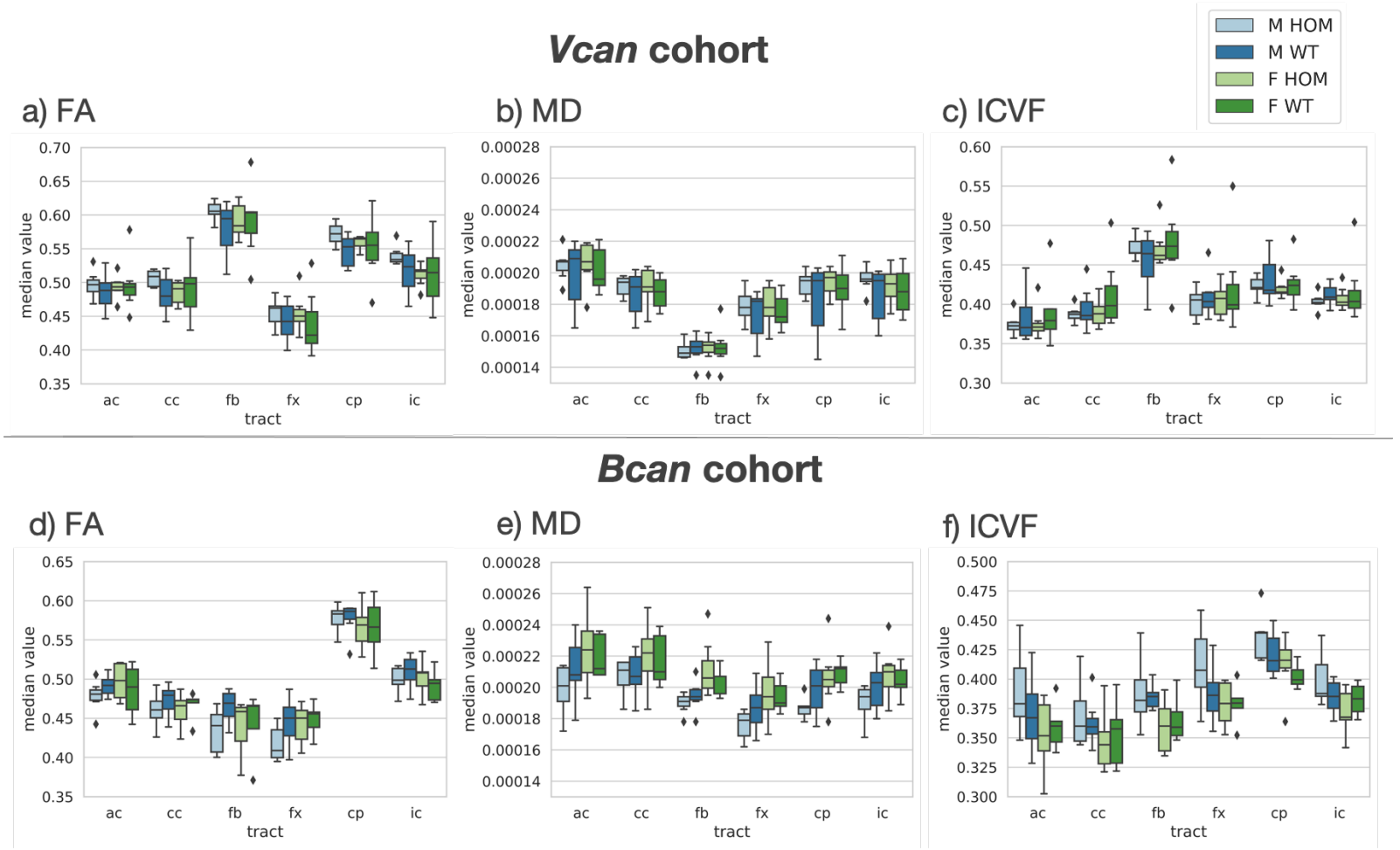

Figure 6.4. Median FA, MD, ICVF in the *Vcan* (a-c) and *Bcan* (d-f) cohorts across the tracts of interest. ac: anterior commissure; cc: corpus callosum; fb: fimbria; fx: fornix; cp: cerebral peduncle; ic: internal capsule. Boxplot components: edges – inter-quartile range, middle line – median value, whiskers – data range (reaching to ± 1.5 beyond the first and third quartiles, respectively), flier points – outliers past the end of the whiskers. Plots produced with Matplotlib v3.3.4.

6.3.1.4. R_2^* and QSM

Statistically-significant differences in R_2^* were observed in the fimbria (p=0.03) and cerebral peduncle (p=0.02) of *Bcan* males, with lower values in HOMs compared to WTs (**Figure 6.5** d). Near-significant differences were observed in R_2^* in the fornix (p=0.065) (**Figure 6.5** d, lower values in HOMs compared to WTs) and in QSM in fimbria (p=0.062) (**Figure 6.5** e, higher values in HOMs compared to WTs) in *Bcan* males. Lower R_2^* values were observed in male HOMs across all tracts.

Opposite, though non-significant, trends compared to the *Bcan* cohort were detected in the *Vcan* cohort, with consistently higher R_2^* values in male HOMs compared to WTs (**Figures 6.4** d and a, respectively). For R_2^*, the Vcan females displayed opposite trends compared to *Vcan* males, with higher values in WTs compared to HOMs.

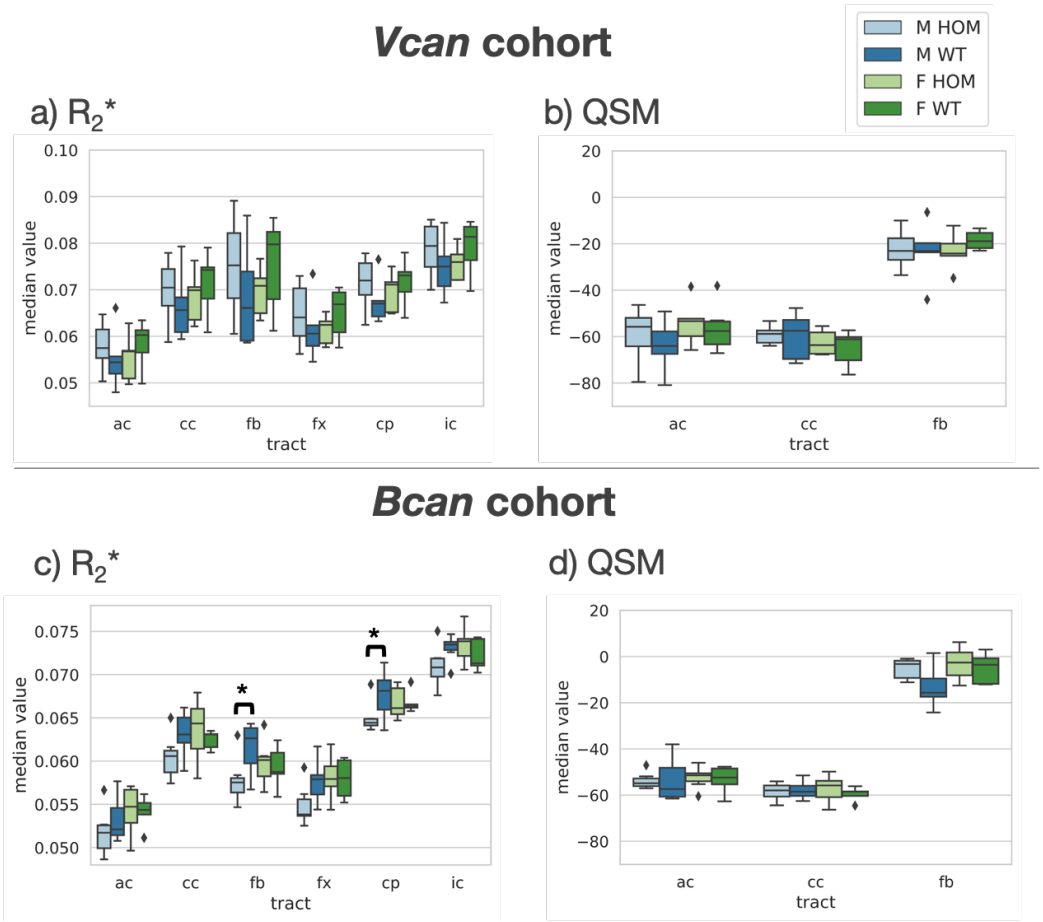

Figure 6.5. Median R_2^* in *Vcan* (a-b) and *Bcan* (c-d) cohorts across the tracts of interest. ac: anterior commissure; cc: corpus callosum; fb: fimbria; fx: fornix; cp: cerebral peduncle; ic: internal capsule. Boxplot components: edges – inter-quartile range, middle line – median value, whiskers – data range (reaching to $\pm$ 1.5 beyond the first and third quartiles, respectively), flier points – outliers past the end of the whiskers. Plots produced with Matplotlib v3.3.4.

6.3.2. MRI-histology analysis

6.3.2.1. Qualitative assessment of the histology data

We conducted a visual comparison across equivalent white matter areas in all samples. No clear differences in staining intensity were observed in PLP, NF or pan-versican by visual assessment of the stained slides in either the *Bcan* or *Vcan* cohorts. **Figures 6.6-6.8** show zoomed-in inserts in males from the *Bcan* and *Vcan* cohorts in the fimbria. The fimbria was chosen as it is the tract where the largest differences across MRI metrics were detected in both cohorts (**Figures 6.4-6.5**). We observe no visible changes in versican V1 either. There is no versican V1 staining in white matter of Vcan HOMs or WTs. Example V1-stained slides are shown in **Chapter 5**. Finally, as reported in **Chapter 5**, *Bcan* homozygous mice show no positive brevican staining.

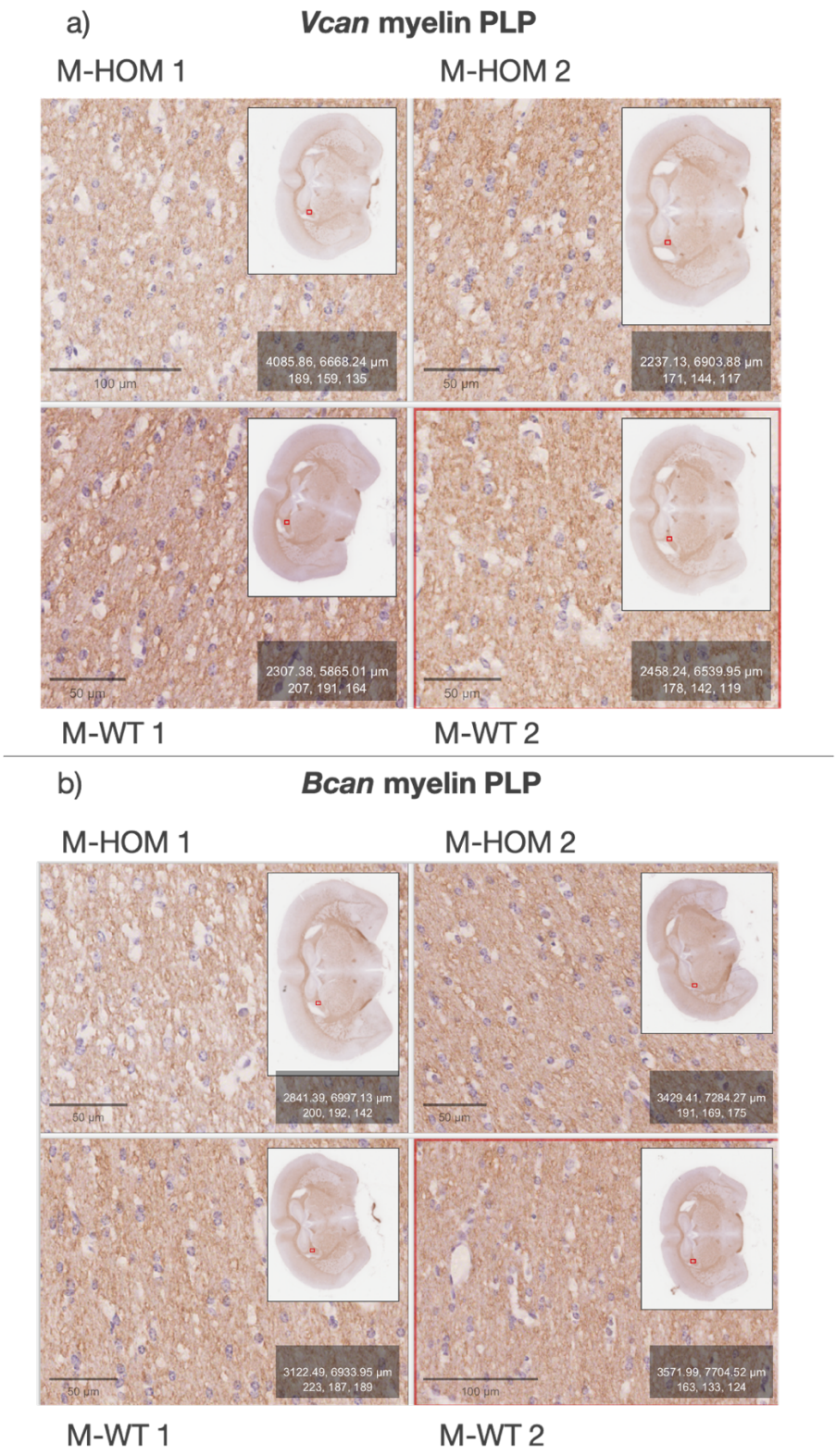

Figure 6.6. Myelin PLP data from *Vcan* (a) and *Bcan* (b) males. No obvious differences between HOMs and WTs are visible between HOMs and WTs. 1 and 2 represent the two mouse brain samples belonging to the same group that were processed for histology. Screenshots taken using QuPath v0.2.3.

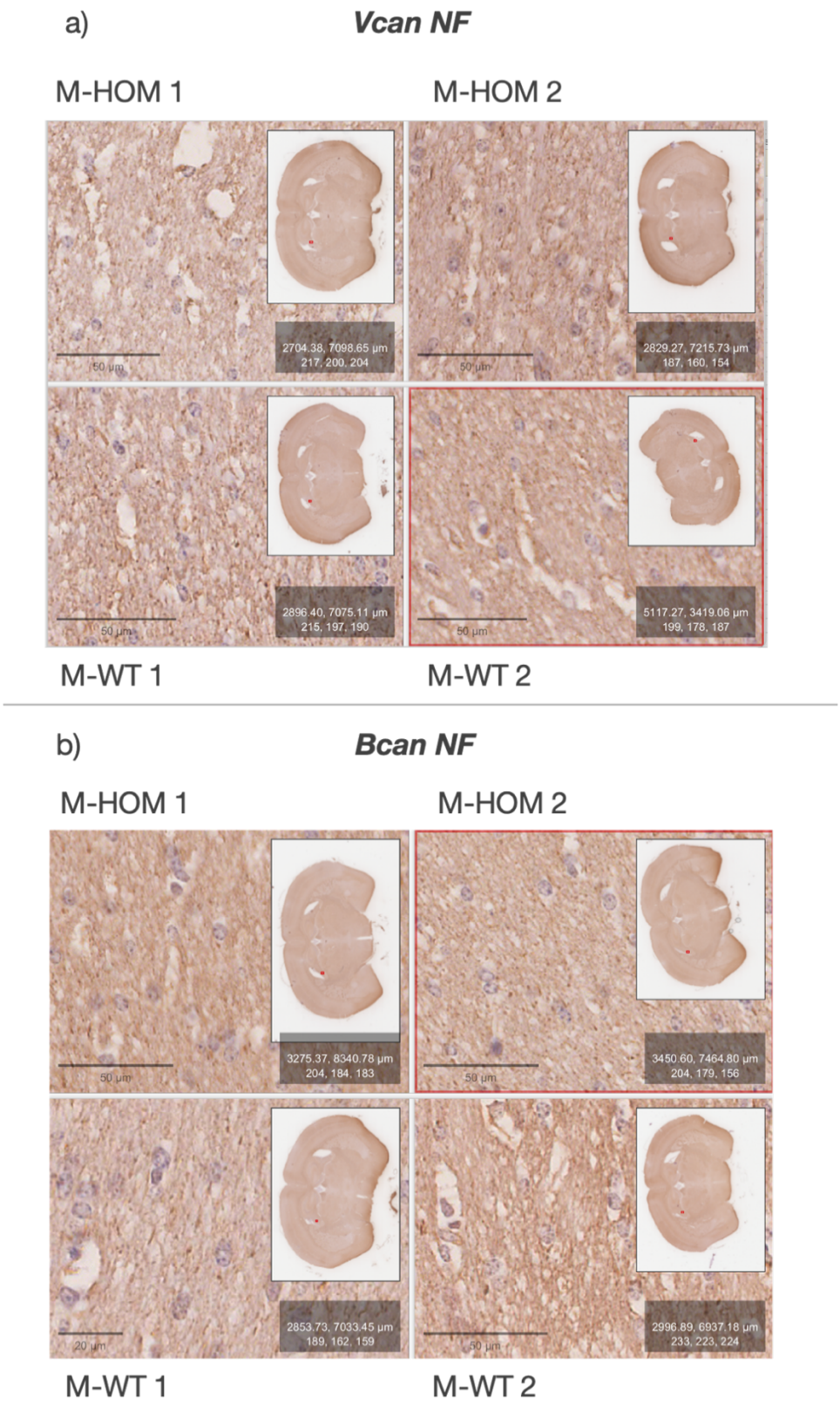

Figure 6.7. NF data from *Vcan* (a) and *Bcan* (b) males. No obvious differences between HOMs and WTs are visible. 1 and 2 represent the two mouse brain samples belonging to the same group that were processed for histology. Screenshots taken using QuPath v0.2.3.

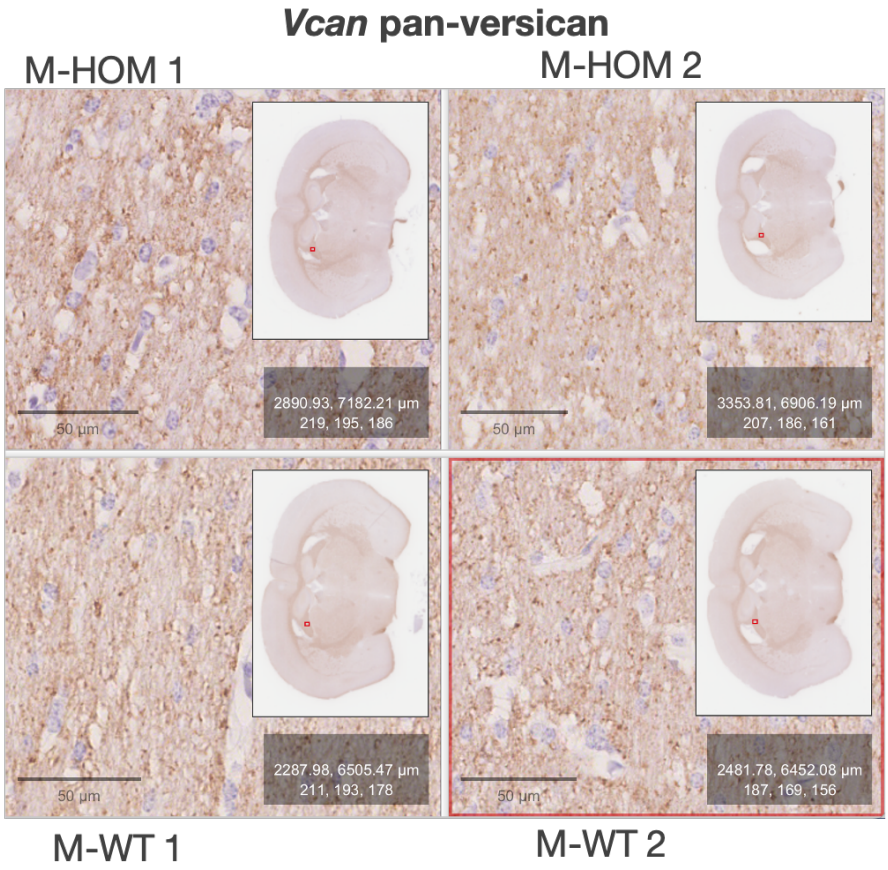

Figure 6.8. Pan-versican data from *Vcan* males. No obvious differences between HOMs and WTs are visible. 1 and 2 represent the two mouse brain samples belonging to the same group that were processed for histology. Screenshots taken using QuPath v0.2.3.

SAF maps also do not reveal any major differences between HOMs and WTs. When pooling together the voxel-averaged SAF values across all locations (A, B, C) for each group (M HOM, M WT, F HOM and F WT), some emerging trends were observed. In particular, higher NF SAF values were seen in male *Bcan* HOMs compared to WTs across all tracts (**Figure 6.9**). 3/4 tracts also display lower PLP SAF values in *Vcan* male HOMs compared to WTs. The pan-versican SAF values do not reveal any obvious trends. Note that this data has not been statistically analysed.

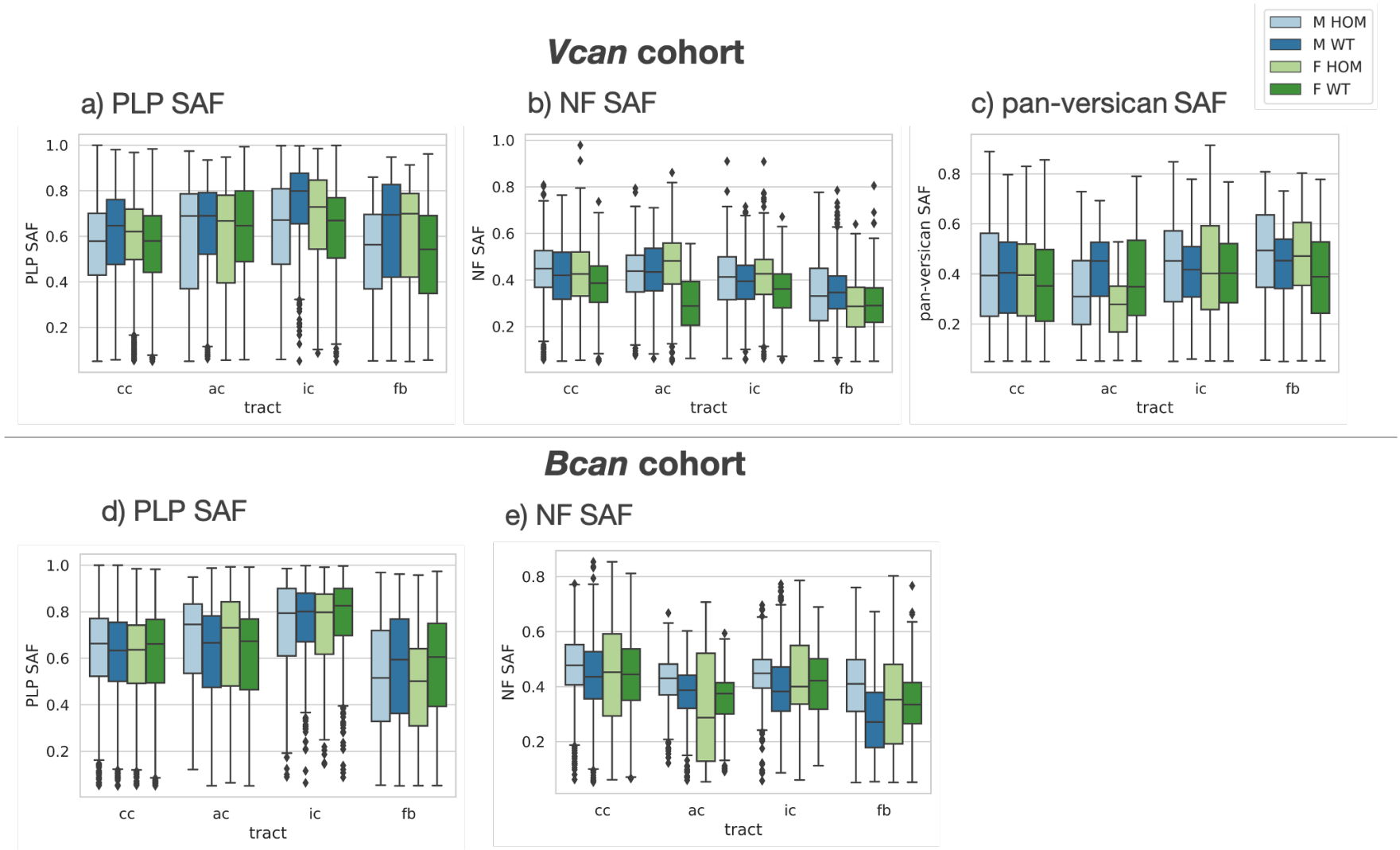

Figure 6.9. Boxplots of MRI-voxel averaged SAF. For each boxplot, the data from all mice in each group (M HOM, M WT, F HOM, F WT) is pooled together across all histological slices (at locations A, B, C), for each tract. Plots produced with Matplotlib v3.3.4.

6.3.2.2. MRI-histology voxel-wise correlations

We observe strong correlations of PLP with all five MRI metrics (FA, MD, ICVF, R_2^* and QSM, **Figure 6.10**), and weaker correlations of NF with the five MRI metrics (**Figure 6.11**), although following the same trends as PLP. The same trends hold for both WTs and HOMs (for PLP, NF and pan-versican).

Interestingly, we observe strong correlations between pan-versican and brevican and the four MRI metrics. Pan-versican shows the same direction of correlation as PLP for all MRI metrics (**Figure 6.12**). Brevican shows an opposite direction of correlation (**Figure 6.13**) compared to PLP, NF and pan-versican for all MRI metrics.

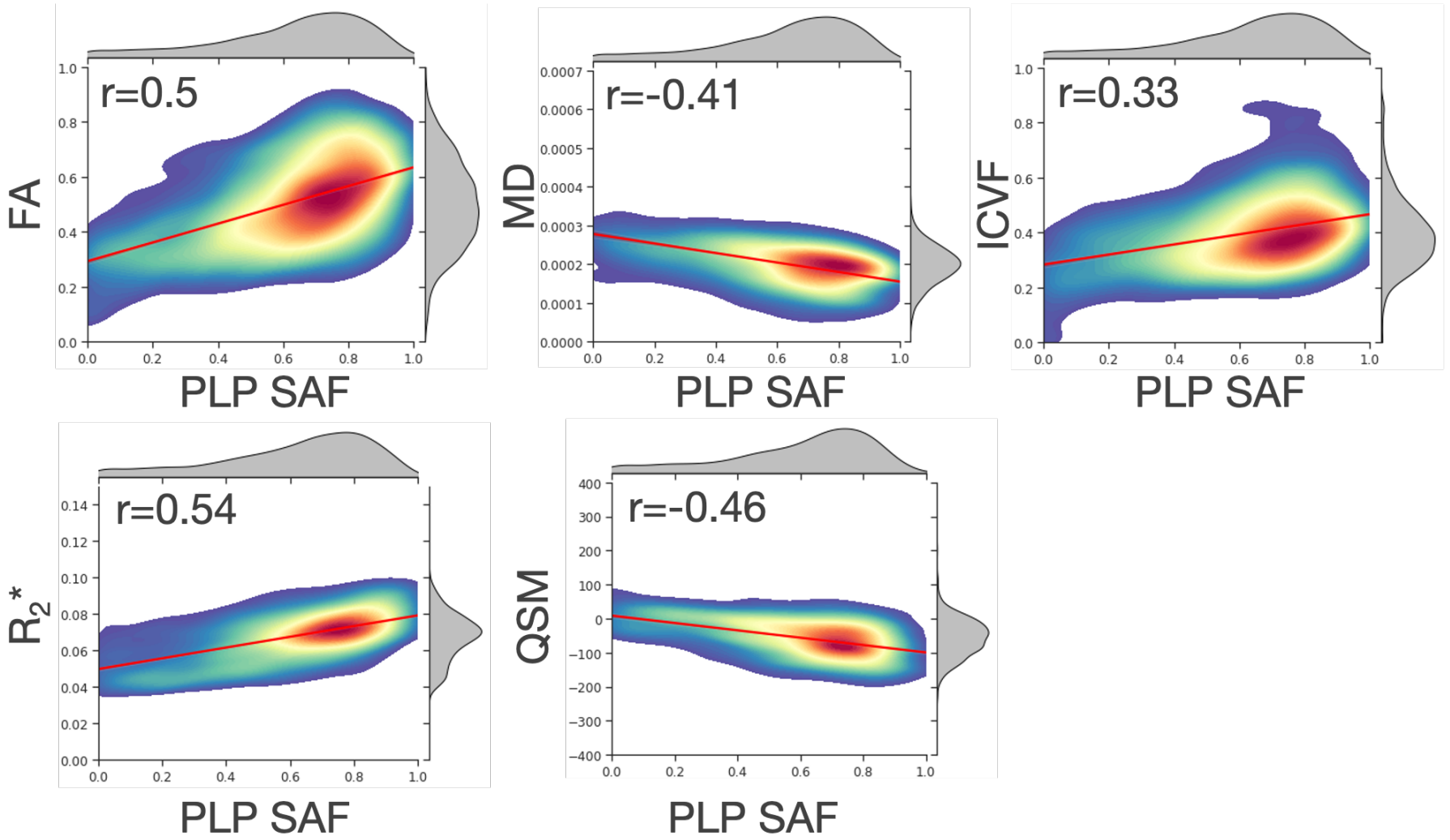

Figure 6.10. PLP SAF-MRI correlations at the MRI voxel level. Data from all white matter tracts across the three ROIs and all wild-type mice from the *Bcan* cohort. r: Pearson correlation coefficient. Plots produced with Matplotlib v3.3.4.

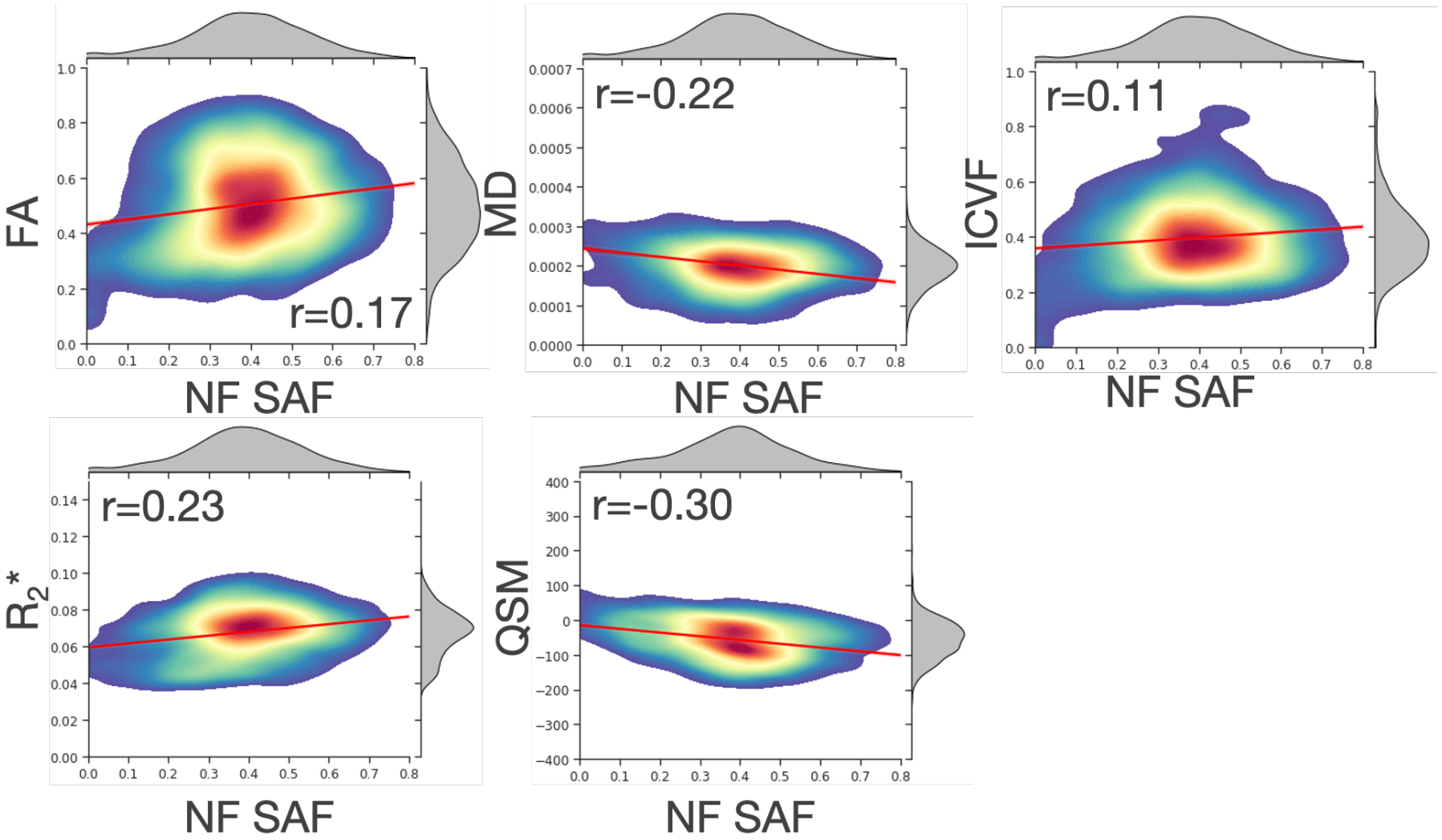

Figure 6.11. NF SAF-MRI correlations at the MRI voxel level. Data from all white matter tracts across the three ROIs and all wild-type mice from the *Bcan* cohort. r: Pearson correlation coefficient. Plots produced with Matplotlib v3.3.4.

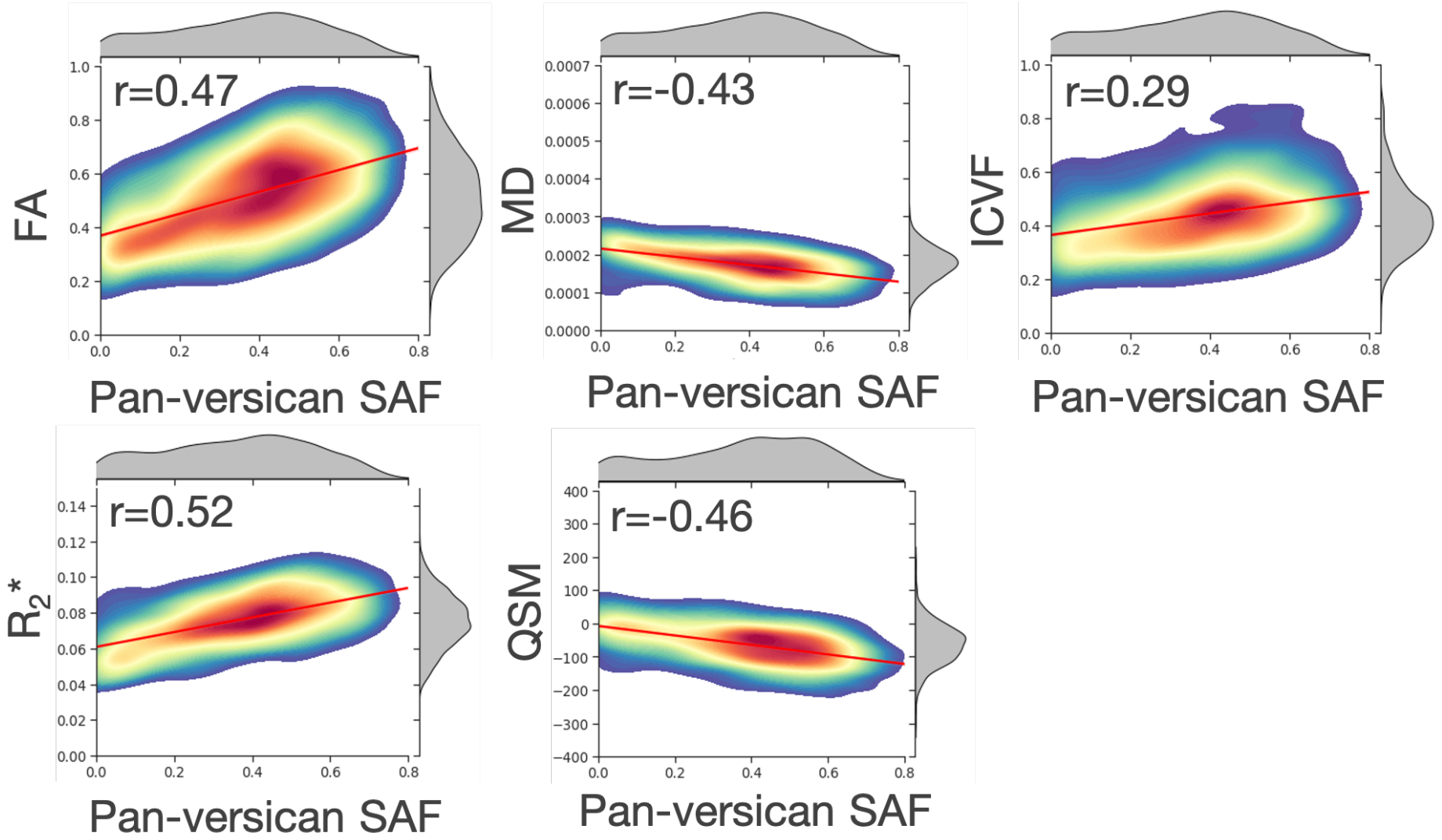

Figure 6.12. Pan-versican SAF-MRI correlations at the MRI voxel level. Data from all white matter tracts across the three ROIs and all wild-type mice from the *Vcan* cohort. r: Pearson correlation coefficient. Plots produced with Matplotlib v3.3.4.

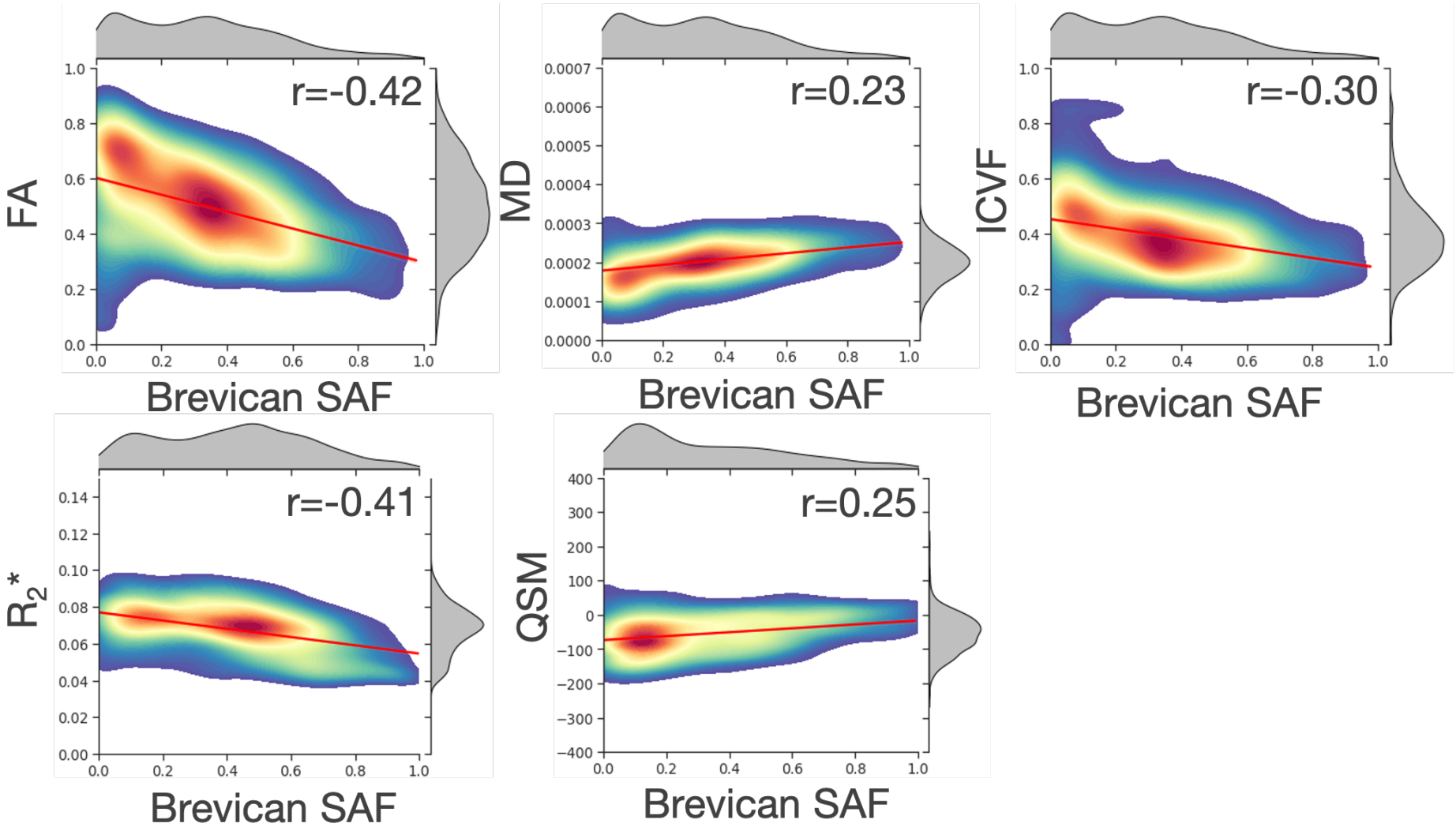

Figure 6.13. Brevican SAF-MRI correlations at the MRI voxel level. Data from all white matter tracts across the three ROIs and all wild-type mice from the *Bcan* cohort. r: Pearson correlation coefficient. Plots produced with Matplotlib v3.3.4.

When evaluating clustering across the above voxel-wise corelations to assess whether there is a separation between WTs and HOMs, no major trends were visible in either of the cohorts. When inspecting correlations across individual tracts, some trends emerge. In *Bcan* males, tracts displaying the greatest difference in R_2^* - the metric with the strongest phenotype from our MRI analysis – also show increased PLP and NF SAF (**Figure 6.14**)

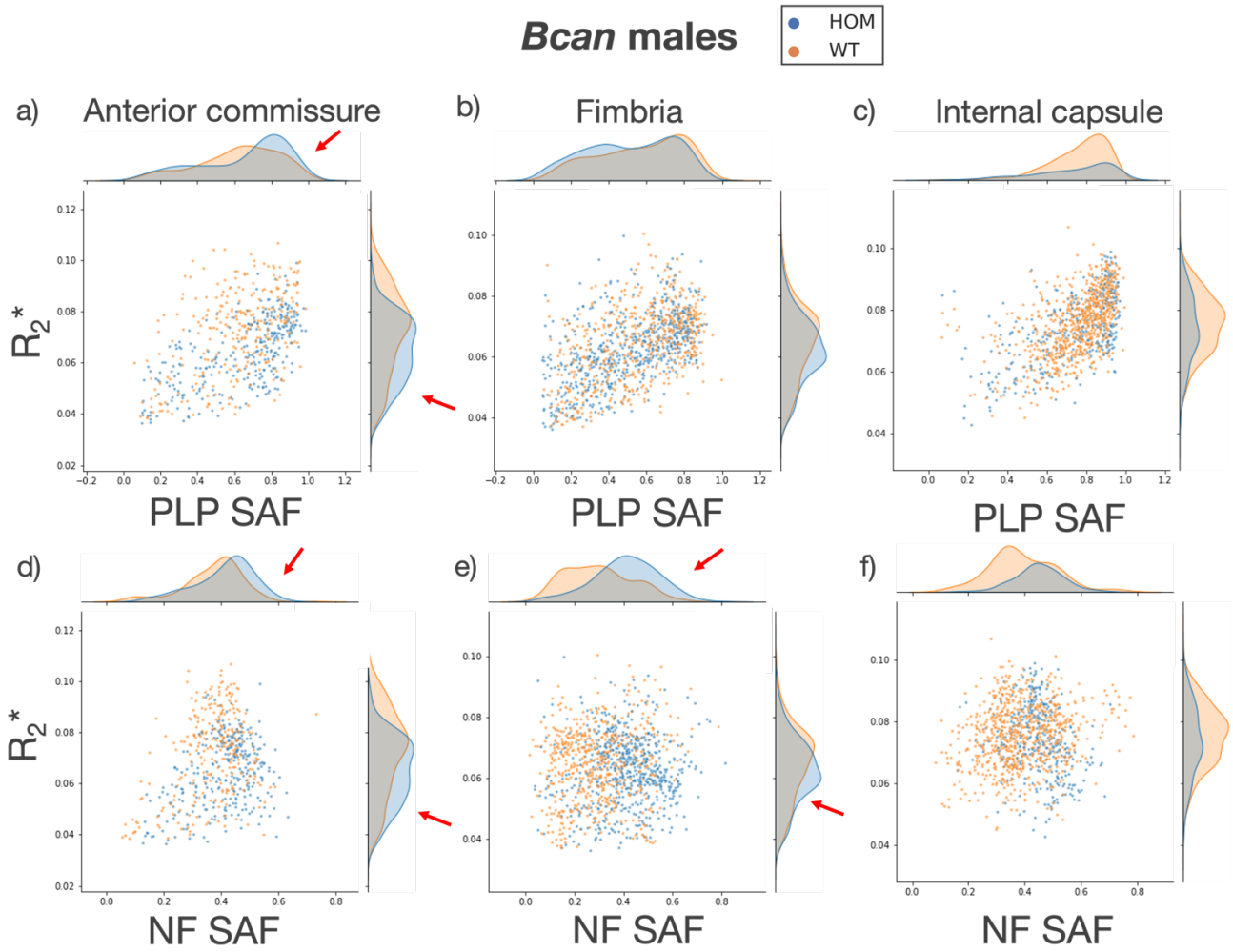

Figure 6.14. Evaluating clustering patterns in R_2^* and PLP/NF scatterplots across various tracts in *Bcan* males (2 HOMs and 2 WTs). The red arrows indicate lower myelin PLP/NF SAF and corresponding lower R_2^* values in some of the tracts. Plots produced with Matplotlib v3.3.4.

6.3.2.3. Composite modelling

The results obtained from fitting the five β coefficients using the composite model are presented in **Table 6.3**. The estimated total fractional genetic contribution to the MRI metrics ranges from 0.113 to 0.266. The genetic contributions to myelin PLP, NF and the MRI metrics indicate that the genetic changes that lead to changes in MRI metrics are mediated by the histological features we measured, since the β coefficients for those pathways are not trivially small compared to β_{g,MRI_metric}.

Cohort	MRI metric	β_{g,MRI_metric}	$\beta_{g,PLP}$	β_{PLP,MRI_metric}	$\beta_{g,NF}$	β_{NF,MRI_metric}	Total frac. genetic contrib. to each MRI metric	Frac. genetic contrib. to PLP	Frac. genetic contrib. to NF
Bcan	FA	-0.09	-0.071	0.544	-0.106	-0.016	0.266	0.113	0.178
	MD	4e-6		-1.5e-4		-1.6e-4	0.113		
	ICVF	-0.059		0.321		0.026	0.222		
	R_2*	-0.003	-0.074	0.029	-0.102	-0.013	0.160	0.116	0.181
	QSM	27.766	0.052	-129.806	-0.120	44.022	0.137	0.084	0.145

Table 6.3. *Bcan* cohort composite modelling results. Voxel-wise data from the fimbria (location B) in *Bcan* males was included in the analysis. β_{g,MRI_metric}, $\beta_{g,PLP}$, β_{PLP,MRI_metric}, $\beta_{g,NF}$, β_{NF,MRI_metric}: fitted parameters of the composite model. Frac. genetic contrib.: fractional genetic contribution. Note that the $\beta_{g,PLP}$ and $\beta_{g,NF}$ differ for each MRI modality as white matter masks in each modality's native space (and a reduced mask that only includes voxels within 30° of the plane perpendicular to B_0 for QSM) were resampled directly into the histology native space.

6.4. Discussion

In this chapter, we describe a reverse translational approach to interpret findings from the UK Biobank neuroimaging GWAS using rodent models.

Our *ex-vivo* multi-modal MRI results confirmed our hypothesis that alterations in the *Vcan* and *Bcan* genes in mouse models lead to similar phenotypes to those observed in UK Biobank GWAS, in either FA, MD, ICVF, R_2^* or QSM. In the *Bcan* mouse model, we observed clear differences across most MRI metrics in males, but not females. We found statistically-significant differences in R_2^* and grey-white matter contrast, with lower values in HOMs compared to WTs. Male HOMs also presented lower FA, MD and higher ICVF.

Based solely on the MRI results, we can hypothesise what might lead to the detected signal changes. It is known that brevican inhibits neurite outgrowth and myelination under certain conditions. Its absence in the *Bcan* HOMs can lead to increased myelination and neurite density, as presented earlier. If this is associated with increases in fibre dispersion, this could explain the reduction in FA and MD and increase in ICVF. R_2^* reductions are usually associated with myelin breakdown and R_2^* correlates positively with myelin stains (**Figure 6.10**) and NF (**Figure 6.11**). However, in our MRI-histology correlations, we observe lower R_2^* values in HOMs associating with higher myelin PLP and NF values (**Figure 6.14**, a and d), as well as higher NF SAF across tracts in *Bcan* HOMs (**Figure 6.9**). This can also be caused by increased fibre dispersion. The dependence of R_2^* on fibre orientation with respect to B_0 has been studied previously (Lee *et al.*, 2011), which suggests that changes in fibre dispersion can possibly also affect the R_2^* measurements. The reduction in grey-white matter ratio, if due only to changes in white matter, indicates a decrease in R_2, likely with the same biological origin as the observed reduction in R_2^*.

Opposite trends are observed in the *Vcan* mouse model, with higher FA and R_2^* values in HOMs compared to WTs. This suggest that an opposite biological mechanism is present in the *Vcan* mice. While *Bcan* mice lack brevican, the consequence of the

genetic modification in the *Vcan* mice is unclear. We do not observe any versican V1 in white matter in either HOMs or WTs. The genetic modification present in *Vcan* HOM mice may lead to compensatory mechanisms such as accumulations or reductions in the other isoforms (particularly V2, present only in the brain). However, we do not observe any clear changes in pan-versican (essentially a versican V2 stain) in the *Vcan* HOM mice. As brevican and versican belong to the same class of proteins and are expected to have similar roles, it is likely that their absence vs. accumulation will lead to opposite effects. As the current histology data does not show any changes in versican V2, it is possible that these changes are either very subtle or that a yet unknown biological mechanism is causing the changes detected using MRI.

No volumetric differences were identified across the four groups of either cohort, indicating that the observed differences are not due to partial volume effects.

The absence of a phenotype in females and lack of statistically-significant results in the *Vcan* cohort can be possibly explained by the limited statistical power. This is likely both due to working with a reduced cohort size than initially planned, as well as an underestimation of the expected effect size when estimating the initial cohort size (7 mice/group). Only 26/28 *Bcan* mouse brain samples were included in the MRI analysis (2 F WT missing from the cohort), 28/28 Vcan samples were included in the dMRI analysis but only 22 samples were available for the QSM/R_2^* analysis (1 M HOM, 2 M WT, 2F HOM, 1F WT missing from the full cohort). Power calculations using the acquired data estimate that we need about 10 samples/group to detect statistically-significant differences.

Sex differences in rodent brains have been observed under many experimental conditions, motivating the need to include both sexes in studies (Premachandran, Zhao and Arruda-Carvalho, 2020). Despite this, female rodents are often avoided in experiments due to concerns that their estrous cycle leads to increases in the variance of the measured traits (Mauvais-Jarvis, Arnold and Reue, 2017). This can be problematic, preventing the discovery of insightful sexual differences. Interestingly, sexual differences in the ECM have been previously reported in literature (Batzdorf *et*

al., 2022). Here, researchers have showed changes in brain elasticity associated with ECM content that differ between male and female mice in both healthy and inflamed brains. The differential ECM content was also reflected in sex-specific gene expression patterns. Based on these findings, there is a likelihood of a biological sex difference in ECM which is reflected in our results. Another possibility is that variance in the measured traits is higher in females compared to males and a larger sample size is needed to identify equivalent effects.

Our preliminary histological results were unable to detect major differences in myelin PLP and NF in the *Bcan* and *Vcan* cohorts or pan-versican in the *Vcan* cohort. However, there is a trend of higher NF in male HOM *Bcan* mice compared to WTs, which is consistent with our MRI results. Due to the small sample size (2 samples/group), we are underpowered when conducting statistical analyses. In **Chapter 7**, we will present our plans for acquiring histology data in all scanned *Vcan* and *Bcan* mice in subsequent work, which will boost statistical power.

The preliminary histology results suggest the following scenarios: Either the differences observed in the MRI metrics are reflected by subtle differences in histology, requiring a larger sample size, or other biological mechanisms influence the detected differences in the MRI metrics. While there is a clear lack of brevican in *Bcan* knock-outs (as shown in **Chapter 5**) we do not detect any clear changes in pan-versican across the tracts of interest. The effect of the genetic modification in our *Vcan* mouse model on white matter microstructure is yet unresolved. These findings agree with reports from literature: no major changes in myelination were detected in a V2 knock-out *Vcan* mouse model (Dours-Zimmermann et al., 2009).

The correlations between PLP and NF and MRI metrics (**Figures 6.11** and **6.12**) agree with those reported in literature (Chandran et al., 2012; Chang et al., 2017; Hametner et al., 2018; Mollink et al., 2019; C. Wang et al., 2020b; Kor et al., 2022). We report previously-undescribed correlations between pan-versican and brevican and our five MRI metrics of interest. While pan-versican SAF shows the same correlation trends

with the MRI metrics as myelin PLP and NF SAF, brevican shows opposite trends. This suggests that versican and brevican might have more complex, differential roles. The significance of the pair-wise MRI-histology correlations (**Figures 6.10**-**6.13**) in understanding the changes detected in the MRI signal needs further investigation. Confusingly, the observed decrease in FA and R_2^*, and increase in ICVF in *Bcan* HOM males compared to male WTs would be consistent with a decrease in myelin PLP and NF, while the reduction in MD would be consistent with an increase in myelin PLP and NF. This is not what we observe in some of our data; for instance, **Figure 6.14** d shows a reduction in R_2^* and an increase in myelin PLP SAF in the anterior commissure in HOM males. Notably, we have not yet conducted any formal statistical analysis of this data; this hypothesis will be tested in subsequent studies. Similarly, brevican SAF shows opposite correlation trends with all MRI metrics compared to myelin PLP, NF and pan-versican: brevican SAF correlates negatively with myelin PLP and NF SAF. This agrees with literature findings, indicating that the absence of brevican could be associated with increase in myelin and neurites.

Finally, the composite model presented in this chapter describes the general approach we plan to take in our MRI-histology analysis. The presented results constitute preliminary analyses. Here, we used voxel-wise data to test the model and did not account for the strong correlation across voxels. In future work, we will model the point spread function to account for this strong correlation. We will also build on this preliminary work to combine models across all tracts and locations. Conducting the analysis in a larger population will also allow us to boost the model's statistical power.

While the composite model revealed some genetic contributions to the MRI metrics, it is not clear whether they are due to genetic differences in myelin PLP or NF. Nevertheless, it is clear that PLP and NF themselves are influenced by genetics. Hence, future work will aim to dissect the relative contributions of myelin PLP and NF to observed genetic influences on MRI metrics. The approach we take will be based on the full composite model, as described in **Chapter 7**.

6.5. Conclusions

In this chapter, we employed a reverse translational approach to investigate the biological contribution to UK Biobank hits using mouse models. We first showed that mouse models of the *Vcan* and *Bcan* genes display related MRI phenotypes to those observed in UK Biobank, with a stronger phenotype present in the *Bcan* mouse model, a brevican knock-out. This pointed to a causal relationship between changes in genes encoding ECM proteins and neuroimaging phenotypes. We followed up with a preliminary histology investigation using a subset of mice from each cohort. While no major changes in myelin PLP or NF were detected, some emerging trends of differences in NF in *Bcan* males could be observed. Given the currently-available data, we cannot conclude how myelin PLP or NF directly contribute to the observed MRI changes. A detailed proposal for follow-up experiments which will be conducted during post-doctoral studies will be highlighted in the final chapter of this thesis.

Chapter 7: Final remarks and future directions

While genome-wide association studies can provide relevant insights into how genes may influence traits, they do not provide the basis to establish a causal relationship between genetic variation and phenotypes. This thesis sets out to explore if there is a causal relationship between genetic variations in the *VCAN* and *BCAN* genes and the associated neuroimaging phenotypes observed in UK Biobank genome-wide association studies. To achieve this, we studied equivalent imaging phenotypes in rodent models with modifications in the *Vcan* and *Bcan* genes.

To facilitate this study, we first developed multi-modal *ex-vivo* MRI acquisition protocols and post-processing pipelines (detailed in **Chapter 4**) – the first implemented protocols on a new Bruker 7T scanning facility. These protocols and pipelines are currently used by other research groups for *ex-vivo* rodent studies in our centre.

We subsequently developed IHC protocols optimised for mouse brain samples prepared for *ex-vivo* MRI, yielding stained slides presenting minimal artifacts. This histology data can be straightforwardly registered to MRI and processed to extract quantitative metrics. As part of this work, we also optimised IHC staining protocols for versican (V1 and V2 isoforms) and brevican. This was detailed in **Chapter 5**.

In **Chapter 6**, we analysed the MRI and histology data of mouse brains from the *Vcan* and *Bcan* cohorts. We demonstrated that these mouse models display equivalent MRI phenotypes to those observed in the UK Biobank cohort. Based on our current histology results, we cannot pinpoint the exact biological mechanisms behind the differences detected with MRI, as we are unable to detect major differences in myelin PLP, neurofilaments (NF) or pan-versican. We did however observe some emerging trends suggesting changes in NF in *Bcan* males that agree with the observed MRI signal changes. This chapter will detail plans for subsequent post-doctoral studies which will build on the findings presented in this thesis.

7.1. Future work

7.1.1. Structure tensor imaging

Our IHC histology data can also be processed using structure tensor imaging (STI) (Budde and Frank, 2012). This provides a complementary quantitative metric, estimating the primary fibre orientation and anisotropy within each pixel and providing information regarding structural organisation of proteins of interest. STI fibre orientation distributions (FODs) can also be derived from a histogram of pixel orientations in a region of interest. STI maps can help investigate whether increased fibre dispersion causes the observed changes in the MRI metrics.

While STI is usually applied on myelin-stained data (Budde and Annese, 2013; Seehaus *et al.*, 2015; Bastiani *et al.*, 2017), recent studies have shown that other stains – e.g., Nissl – can also be used to study the underlying axonal architecture in white matter (Schurr and Mezer, 2021). In white matter, dense white stains – such as myelin PLP, NF but also glial protein stains – can be expected to yield similar FOD or anisotropy estimates as their respective proteins surround axonal fibres (Budde and Frank, 2012).

We have run STI on our histology datasets using an extension to the automatic SAF pipeline (Kor et al., 2022). Pixel-wise estimates of orientation were calculated across a kernel size of 4 x 4 pixels using the pipeline's DAB segmentation output. Directionally-encoded colour (DEC) representative images of the STI outputs for the myelin PLP, NF, brevican and pan-versican stains are shown in **Figure 7.1**. The STI maps of myelin PLP and NF stains followed the expected axonal orientation in white matter, revealing the intricate fibre architecture. The brevican STI maps showed no obvious orientation dependence in either grey or white matter. The pan-versican STI maps revealed similar orientation dependence to myelin PLP and NF in white matter.

In our follow-up studies, we will conduct a quantitative analysis of the FOD across white matter regions of interest in these datasets, seeking to determine whether this

correlates with the observed changes in the MRI metrics. This will help elucidate whether the observed changes in the MRI metrics are indeed due to changes in fibre dispersion.

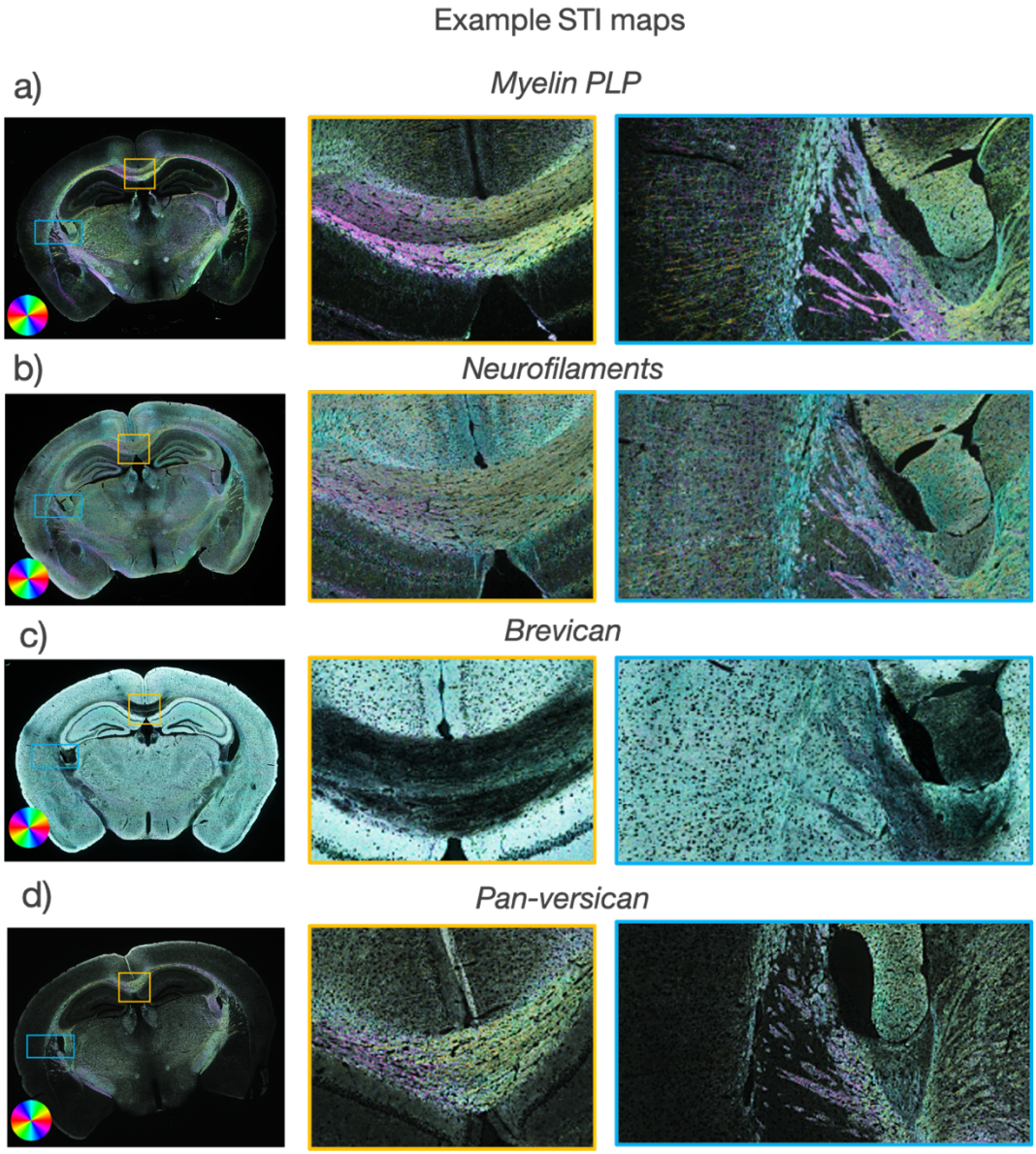

Figure 7.1. Example STI maps. a) Myelin PLP STI map. b) Neurofilaments STI map. c) Brevican STI map. d) Pan-versican STI map.

The pair-wise correlations presented in **Chapter 6**, **Figures 6.10-6.13** also warrant further investigation. In line with recently published work from our group (Kor et al., 2022), we can use partial correlation to account for co-variance between stains (e.g.,

between myelin PLP and NF). Multiple linear regression can also be used to understand the relative contributions of the stains to each of the MR metrics of interest.

7.1.2. Subsequent *ex-vivo* MRI data acquisition

We plan to address the challenge of working with a smaller cohort than initially planned. We conducted a power calculation and estimated we need 10 samples/group to detect statistically-significant differences. We have obtained new *Bcan* and *Vcan* samples from our collaborators at KCL and MRC Harwell, which are going to be scanned using the same protocols presented in this thesis.

7.1.3. Subsequent histology data acquisition

We did not observe any major changes in myelin PLP or NF in our histology dataset containing 2 samples/group for each of the *Vcan* and *Bcan* cohorts. A larger cohort size will boost statistics and help elucidate the respective genetic contributions of myelin PLP and NF to the observed MRI signal changes.

Considering the current results, we propose to use all our available samples in the subsequent histology analysis, and analyse one individual brain region at a time. We will obtain data from coronal slices in the mid-corpus callosum for all samples and stain the slices using myelin PLP, NF, pan-versican and versican V1 (for *Vcan* mice) and brevican (for *Bcan* mice). Time-permitting, we will acquire data from other regions in the brain as well.

7.1.4. Composite modelling

The composite model as depicted in **Chapter 6** separates out the different pathways by which genetics cause changes in a given MRI metric. In future work, we will be able

to estimate how much of the genetic alterations in MRI metrics are mediated by genetically-driven changes in NF or myelin PLP. This will enable us to inquire about the relative contributions of the stains to differences in MRI metrics that are attributable to genetics.

The composite modelling results included in **Chapter 6** were preliminary. Once we will acquire data in all *Vcan* and *Bcan* mouse brain samples scanned for *ex-vivo* MRI, we will work with the median values across each individual tract. The model will be applied to multiple data categories (sex and tract) for each modality and subsequently combined. We are in the process of identifying the correct statistical framework for this mixed-effects model.

7.1.5. Long-term research plans

Prof Miller's group will continue studying mouse models of genes relevant to UK Biobank in subsequent longer-term work, as part of a larger study using biophysical compartmental modelling of the MRI signal to aid the translation of UK Biobank findings to clinical settings. Here, the group will seek to validate the biophysical models they develop in *Vcan* and *Bcan* rodent models.

7.2. Final remarks

The work presented in this thesis sits at the interface between MRI, cross-species neuroscience and genetics. Establishing the presented tools and protocols thus required strong collaborations with researchers across various groups within our centre. Our research outputs have been shared openly and are only just starting to be used by groups abroad. This brings unique value to this work and will hopefully continue to foster fruitful collaborations.

Finally, the findings presented in this thesis pave the road for exciting new research which will hopefully enhance the value of the UK Biobank neuroimaging studies.

Appendix A: A protocol for high-resolution multi-modal MRI phenotyping of *ex-vivo* mouse brains – Data acquisition and processing

Materials and reagents

1. Bottles for storing the samples, with screw cap
2. 15 ml high clarity polypropylene conical falcon tubes, 17x120mm style (Corning science, Mexico, catalog number 352096)
3. Fluorinert (3M, catelog number: FC-770)
4. Gadovist (Gadolinium-based contrast agent; Bayer, catalog number: 30351004)
5. Heparin 10 000USP/10ml (Wockhardt, catalog number: 30170145)
6. Ketamidor 100mg/ml solution for injection (Ketamine; Chanelle Pharma, catalog number: 30395940)
7. Phosphate-Buffered Saline solutiin (Corning, catalog number: 21-040-CV)
8. 16% pre-made paraformaldehyde (PFA) 10 x 10ml (Electron Microscopy Services, catalog number: 15710)
9. Rompun 25ml (Xylazine, Bayer, catalog number: 2150569)
10. Sodium azide (Sigma Aldrich, catalog number: 102141474)
11. Materials for perfusion: 0.5ml insulin syringe, surgical scissors, forceps, haemostat (1/mouse), absorptive pads, cotton gauze, masking tape, surgical tape, butterfly needle, vacutainers, markers for labelling samples (or label printer).

Experimental models

1. Mouse, Bcan knockout (BCAN-DEL465-EM1-B6N-IC) and wild-type littermates, 10 weeks, males and females (MRC Harwell)
2. Mouse, Vcan point mutation (VCAN-E441A-EM2-B6-IC) and wild-type littermates, 10 weeks, males and females (MRC Harwell)

Equipment

1. 7T Preclinical MRI Scaner, Bruker BioSpec, 70/20 USR (Bruker Biosystems, Etlingen, Germany)
2. 4-channel receive-only cryoprobe (Bruker Biosystems, Etlingen, Germany)
3. 86mm inner diameter volume transmit coil (Bruker Biosystems, Etlingen, Germany)
4. Cryoprobe cradle (Bruker Biosystems, Etlingen, Germany)
5. Custom, 3D-printed mouse brain holders (brains in the skull)
6. Custom, 3D-printed falcon tube holders (for the cryoprobe cradle)
7. Fridge for storing brain samples long-term at 4°C (Liebherr Mediline)
8. Mock cryoprobe for testing sample position (Bruker Biosystems, Etlingen, Germany)
9. Transcardiac perfusion peristaltic pump
10. Tools for performing transcardiac perfusion
11. Vacuum pump and vacuum chamber (Vacuubrand GMBH, Wertheim, Germany)

Software

1. FMRIB Software Library (FSL) (University of Oxford, RRID:SCR_002823 https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
2. FSLEyes (University of Oxford, https://git.fmrib.ox.ac.uk/fsl/fsleyes/fsleyes)
3. MATLAB R2019a or later (www.mathworks.com)
4. MATLAB read2dseq (function for reading Bruker 2dseq files, https://www.mathworks.com/matlabcentral/fileexchange/69177-read_2dseq-quickly-reads-bruker-s-2dseq-mri-images)
5. MATLAB "Tools for NIfTI and ANALYZE image" Toolbox (https://www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image)

6. MATLAB "STI Suite" Toolbox (https://people.eecs.berkeley.edu/~chunlei.liu/software.html)
7. mrdegibbs3D (https://github.com/jdtournier/mrdegibbs3D)
8. Paravision (PV 360, Bruker Biosystems, Etlingen, Germany)

Procedure

A. Perfusion of mouse brain samples and long-term storage

The detailed prefusion protocol is described elsewhere (Cahill *et al.*, 2012). A brief description is highlighted below.

1. Prepare the perfusion setup at a perfusion table, including perfusion pump and stand, absorbent mats and equipment.
2. Anaesthetise the mice with ketamine (150mg/kg) and xylazine (10mg/kg), mixing with a saline solution to have 0.1ml/10g. This should be done via intraperitoneal injection.
3. Open the chest cavity.
4. Clamp the xiphoid process with a haemostat.
5. Insert a butterfly needle into the left ventricle of the heart, cutting the right auricle to allow for blood draining.
6. Using the pump, perfuse the mouse with 30ml of room-temperature perfusion solution (1 x PBS, 1μl/ml heparin and 2mM Gadovist, flow rate: 1ml/min).
7. Follow this up with the second 30ml perfusion solution (4% PFA and 2mM Gadovist, flow rate: 1ml/min) when there is almost no perfusion solution 1 left in the channel.
8. Check that the body is still by flicking the tail.
9. Following perfusion fixation, remove head and trim tissue surrounding the skull. Do not remove the premaxilla and nasal cartilage as this will help stabilise the skull during scanning.
10. Leave in PFA solution (4% PFA, 2mM Gadovist) at 4°C overnight (12h).
11. Transfer the samples to a storage solution (1 x PBS, 0.02% sodium azide, 2mM Gadovist) the following day.
12. Store the samples for long periods of time at 4°C.

Note: The samples should be scanned at the earliest one month after perfusion and ideally before 6 months (de Guzman *et al.*, 2016), to minimise the effect of morphometric changes in the brain that occur during storage and fixation.

B. Sample preparation for scanning and testing

1. Take one sample at a time out of the storge solution.
2. Wipe the skull gently with a tissue to remove any PBS solution residue, otherwise this will show up as hyperintensities on the MRI data.
3. Insert the sample in a 15ml, labelled falcon tube filled with Fluorinert (MR-invisible non-protonated solution) using a 3D-printed holder which allows to keep it in place and at consistent orientations.
4. Vacuum pump for 30min-1h (pressure 10-20mbar) to remove any air bubbles present between the brain and the skull.
5. Place the bubble trap between the holder and the top of the falcon tube, fill the falcon tube with Fluorinert to the top and screw the cap.
6. Ensure that the air bubbles are trapped behind the bubble trap.
7. Transport the sample tube to the scanner, avoiding air bubbles from escaping the trap.
8. Place the sample tube on the custom 3D-printed falcon tube holder.
9. Use the mock cryoprobe to evaluate sample location with respect to the cryoprobe and scanner isocentre to maximise sensitivity.
10. Once the optimal position is determined, fix the falcon tube onto the falcon tube holder with tape.
11. Fix the falcon tube holder to the cryoprobe cradle with tape.
12. Slide the sample into the scanner.
13. Run a Bruker Localiser protocol to evaluate sample location. You will be able to visualise 3 orthogonal planes through the sample. The brain should be roughly in the centre of the image.
14. Perform any millimetric adjustments to the sample location using the cryoprobe bed screw.
15. Re-run the localiser to evaluate whether the location of the sample has improved.

16. Repeat steps 14-15 as many times as necessary.
17. Perform any tuning and matching adjustments needed. The volume coil needs to be tuned and matched manually, whereas the cryoprobe can be tuned and matched using the ParaVision software.
18. Run a series of short acquisitions to evaluate sample integrity. Details of these are presented in **Chapter 4**, **Table 4.3.**
19. Inspect the outputs of the T2-weighted structural scan. It is expected that white matter has lower signal compared to grey matter. A reverse in T2 contrast is indicative of poor perfusion and a faulty sample.
20. Inspect the T2 mapping scans to confirm contrast inversion and evaluate the T2 values across the brain. Generally, poorly-perfused samples have higher T2 values compared to good samples.
21. Inspect the multi-echo GRE (particularly the later echoes) to evaluate the presence of the bubbles. If the bubbles are too big and cover relevant regions-of-interest, we recommend that the samples are returned to the vacuum pump for another 30min-1h session.
22. If needed, repeat steps B5-B21 from this section (minus the T2-weighted and T2-mapping scans) to ensure that the samples are ready to be scanned.
23. The samples can be stored in the falcon tubes in Fluorinert at 4°C for a few weeks following this process. Fluorinert does not alter the quality of the samples (Iglesias *et al.*, 2018).

C. *Ex-vivo* multi-modal MRI data acquisition

Before scanning

1. Remove the sample from the fridge at least 1h before scanning, to allow it to reach room temperature. Temperature fluctuations in the sample can lead to errors in some MRI-derived measures, such as diffusivity estimates.
2. Set the cryoprobe heating window temperature to room temperature (approximately 22°C).

3. Follow steps B6-B22 from the previous section.
4. Acquire a B0 map.
5. Acquire the structural T_2-weighted, dMRI and multi-echo GRE data following the key parameter details and pre-scan adjustments highlighted in **Chapter 4**, **Table 4.2**. Use the MAPSHIM utility (or equivalent) to calculate optimum shim values within a graphically selected volume based on the B0 map. The T_2-weighted data is acquired at 60μm isotropic resolution, while the dMRI and multi-echo GRE datasets are acquired at 100μm isotropic resolution.

   For each sequence, we also performed the following modifications to reduce the presence of aritfacts. These may be optional, depending on your particularities of your hardware setup.

   i. T_2-weighted MRI: We found that modifying the default spoiling parameters to Auto Repetition Spoiler between echoes: OFF, spoiling: 1 cycle and BW=60kHz significantly reduced blurring present in the images. We attributed the blurring to large phase offsets present across echoes.
   ii. dMRI: We empirically determined that using a 1.5 excited slab/FoV ratio for our excitation FoV eliminated an inhomogeneous drop in signal intensity observed on the images (described in **Chapter 4**). We incorporated this modification into the pulse program sheet, which allows local users to adjust this ratio.
   iii. multi-echo GRE: We reduced severe blurring and misalignment across volumes by using a prospective navigator (excitation and readout), applied before each excitation, as well as by increasing the number of dummy scans before each scan to the maximum value, to allow for the temperature to stabilise. Temperature fluctuations occur during this sequence due to the slow heating of the passive shim, which can subsequently heat up the gradient set. We recommend that the multi-echo GRE data is acquired last to prevent this issue from impacting subsequent data acquisitions.
6. Once acquired, perform a quality-control check of the data. Ensure that no major artifacts are present across the imaging volumes for each modality. Inspect the

signal profile across volumes for various voxels around the brain for both the dMRI and multi-echo MGE datasets. For the dMRI data, ensure that the b=0 volumes have roughly similar signal intensity. For the multi-echo MGE data, check that the signal decay curve is roughly exponential. Compare the signal intensity across imaging volumes to previously acquired scans.

Data processing

A. Structural (T_2-weighted) MRI

The T_2-weighted data is used for whole-brain atlas-based segmentation and iterative registration to a cohort-specific standard space. The pipeline for processing the structural datasets is available online (https://github.com/Mouse-Imaging-Centre/pydpiper/tree/main) and extensively described in other papers (Lerch, Sled and Henkelman, 2010; Chakravarty *et al.*, 2013; Friedel *et al.*, 2014).

B. Diffusion-weighted MRI

1. (Optional) If the raw nifti is outputted by the scanner using wrong axis labels (superior-inferior: SI, anterior-posterior: AP, left-right: LR), then the data needs to be reoriented so that the axis labels are anatomically correct. This is done using FSLSWAPDIM.
2. (Optional) If step 1 was performed, reorient the b-vectors to account for the new labels.
3. Perform 3D Gibbs-ringing correction (Step A.1. from this section).
4. Use the first b=0 volume ("Blip Up") and a volume acquired with a reverse phase-encoding direction ("Blip Down") to run FSL's TOPUP (Andersson, Skare and Ashburner, 2003; Smith et al., 2004) and estimate any susceptibility-induced distortions present in the image. Adjust TOPUP's acquisition file input (--datain) accordingly, specifying the phase encoding direction of each volume within the dMRI acquisition. An adjusted TOPUP configuration file (--config) is needed which accounts for the high resolution of mouse MRI data, by modifying the following parameter values: "Resolution (knot-spacing) of warps in mm" to "--warpres=2,1.6,1.4,1.2,1.0,0.6,0.4,0.4,0.2" and "FWHM of gaussian smoothing" to "--fwhm=0.8,0.6,0.4,0.3,0.3,0.2,0.1,0,0". Note that topup does not accept inputs where the phase encoding direction is in the superior-inferior direction as it expects this direction to be allocated to the readout direction (as per human

imaging protocols). The volumes need to be reoriented using FSLSWAPDIM to allow TOPUP to run.

5. The output from TOPUP is fed into FSL's EDDY (Andersson and Sotiropoulos, 2016), which corrects these distortions and any misalignment between volumes which may have occurred due to B_0 drift.
6. Reorient the data into standard anatomical orientation following distortion correction using FSLSWAPDIM.
7. (Optional) Using one of the b=0 volumes, run the pipeline for iterative registration and atlas-based segmentation (described in section A) to obtain a brain mask for the data.
8. Fit the DTI and DKI models using FSL's DTIFIT function. This will output the relevant parametric maps: fractional anisotropy (FA), mean diffusivity (MD), principal eigenvector (V1), apparent kurtosis (AK) etc.
9. Fit the NODDI model using FSL's cuDIMOT tool (Hernandez-Fernandez *et al.*, 2019). This will output the relevant parametric maps: orientation dispersion (OD), intra-cellular volume fraction (ICVF), isotropic volume fraction (ISOVF). NODDI assumes fixed axial diffusivity, d_{ax}, and isotropic diffusivity, d_{iso}, across all imaging voxels. Literature values are routinely used, but these may not account for particularities of the sample processing, such as the perfusion fixation protocol, which may affect water diffusivity in tissue. We used a data-driven approach, highlighted in **Chapter 4**, to derive an optimised d_{ax} value for our samples, $4.7 \times 10^{-4} mm^2/s$. d_{iso} was set to $10^{-3} mm^2/s$. Include a dot compartment for the NODDI fitting, to account for water molecules trapped by fixative.

Shell scripts run steps 1-7 using a single command. Steps 8 and 9 need to be run separately. The pipeline diagram is present in **Chapter 4**, **Figure 4.9**.

C. Multi-echo gradient echo MRI

1. Data preparation stage. For both pipelines:
    1.1. (Optional) If the raw nifti is outputted by the scanner using wrong axis labels (SI, AP, LR), then the data needs to be reoriented so that the axis labels are anatomically correct. This is done using FSLSWAPDIM.
    1.2. Ensure that all your volumes for all your repetitions are co-registered. Use a simple 6 degrees-of-freedom rigid registration if needed (Jenkinson *et al.*, 2002), and register all volumes to the first volume of the first repetition, for example. Apply the transforms derived using the magnitude data to the real and imaginary components of the complex data, for all channels and repeats.
    1.3. Using one of the volumes acquired at earlier echo times, run the pipeline for iterative registration and atlas-based segmentation (described in section A) to obtain a brain mask for the data. The brain mask is essential for QSM processing as non-brain voxels will affect susceptibility estimates.

2. The R_2^* pipeline:
    2.1. Perform 3D Gibbs-ringing correction (Step A.1. from this section) for all volumes and repeats.
    2.2. Average the repetitions using FSL's FSLMATHS.
    2.3. Perform voxel-wise fitting to estimate R_2^* values for each imaging voxel. A noise parameter should be incorporated into the fitting, and the fitting should assume Rice-distributed data.

3. The QSM pipeline:
    3.1. Perform coil-combination and phase-offset removal across all coil channels. This will avoid the presence of any unwanted artifacts in the subsequent processing stages. We use the COMPSER method, which subtracts the phase at the first echo from all subsequent echoes

(Robinson *et al.*, 2017). The data is coil-combined using a complex sum over all coil channels.

3.2. Obtain a voxel-wise estimate of the field map using the complex data for all echoes. We recommend using the method detailed in Wang et al (C. Wang et al., 2020), which uses the coil-combined complex data for the non-linear fitting. This method also derives a phase offset map.

3.3. Use the phase offset map to remove any unreliable voxels at the edge of the brain presenting large phase values. This will reduce the appearance of artifacts in the susceptibility parametric map.

3.4. Perform background field removal to get rid of contributions to the field map that come from outside the region of interest (in this case, the brain). We use the variable-kernel Sophisticated Harmonic Artifact Reduction for Phase Data (vSHARP) algorithm (Li, Wu and Liu, 2011; Schweser *et al.*, 2011). The vSHARP algorithm should be run exclusively for voxels within the brain mask. We used a kernel size of 2mm as an additional input to vSHARP.

3.5. Perform a dipole inversion algorithm to convert the estimated field contributions from the tissue of interest to voxel-wise QSM maps. We use the iterative LSQR (iLSQR) algorithm (Li *et al.*, 2015). The field direction needs to be specified when running this stage. If the setup used is equivalent to ours, then the field direction should be roughly aligned with the anatomical and image anterior-posterior (AP) axis.

3.6. Reference the susceptibility values to the whole-brain by subtracting the mean voxel value across the whole brain from each individual voxel.

Steps 1.1.-1.3. are run using a single command, which prepares the data for model fitting. Steps 2.1-2.3 and 3.1.-3.6 are run in a separate command. The pipeline diagram is present in **Chapter 4**, **Figure 4.11**.

The pipelines openly available to download on Zenodo or our centre's gitlab page (https://git.fmrib.ox.ac.uk/preclinical-imaging). The scripts contain anonymised file-

paths which will need to be edited to enable running on a cluster facility. The commands for submitting jobs to the cluster should also be edited. The pipelines should be compatible with any other data acquired on a similar Bruker scanner and using an equivalent protocol. The pipelines outputs intermediary files at every stage, which can be inspected following its completion. The repositories also include a guide for evaluating the outputs of the pipeline at all steps, with example outputs (/Pipeline_Guide_and_QC/ Pipeline_Guide_and_QC.md).

Appendix B: Optimisation of the anti-versican and anti-brevican immunohistochemistry staining steps in mouse brains prepared for *ex-vivo* MRI

1. H_2O_2 incubation time

H_2O_2 incubation is a blocking step conducted in IHC staining protocols to block endogenous peroxidase activity, which can react with DAB leading to false positives, or non-specific background staining. This step reduces background staining significantly. Although 10-15min of H_2O_2 incubation is recommended according to literature, we found 30min-1h to be needed sometimes to achieve the best results.

2. Heat-induced epitope retrieval

HIER is required to break methylene bridges formed during fixation and reveal the antibody binding site, boosting the true positive staining. Three common options we tested are: autoclave in citrate buffer (pH6), microwave in citrate buffer (pH6) and autoclave in Tris-EDTA (pH9). We found that different antibodies responded differently to the three HIER methods and therefore ran this optimisation step for each antibody.

3. Chondroitinase ABC treatment

The ChABC enzyme antigen retrieval step is specific to the versican and brevican staining protocols. ChABC is an enzyme which degrades the side chains of the versican and brevican molecules. *In-vivo*, it is used to render these proteins inactive, sometimes even as a potential treatment strategy for neural injury (Mahajan, 2018). This is due to versican and brevican's inhibitory effect on myelin regeneration, which

can be suppressed by ChABC treatment. Interestingly, the same ChABC mechanism can be used *ex-vivo* in IHC experiments to reveal the antibody binding site and enhance stain quality. In our experiments, we tested whether ChABC antigen retrieval was needed and, if so, which of the recommended stock solutions – Tris-HCL or TSA-BSA – and incubation times yielded the best results.

4. Blocking buffer solutions

A blocking buffer solution (or, simply, blocker) needs to be used before primary antibody incubation. This will reduce non-specific binding of the antibodies. A blocker will attach to the potential non-specific binding sites in the sample. Different blockers are available commercially. During our experiments, we also tested which of the available blockers – Dako Protein Block Serum Free or Fetal Calf Serum 10% - performed better. For the optimisation of the versican and brevican stains, we found Dako Protein Block Serum Free to yield significantly better results, with minimal batch-to-batch variations in staining results.

5. Primary antibody incubation time and concentration

Finally, we optimised the antibody concentration. Our initial approach was to test a broad range of concentrations (e.g., 1:100, 1:500, 1:1000, 1:1500, 1:2000, 1:2500) to identify the optimal level for our samples. We narrowed down on the concentration which presented sufficient specific, yet limited non-specific staining (e.g., if 1:500 was optimal, we then tested 1:400, 1:500, 1:600). On a case-by-case basis, depending on results, for instance if the staining was quite faint even at high concentrations, we explored overnight (18h) primary antibody incubation.

During the protocol optimisation, we adhered strictly to the timings of each individual step. To avoid any cross-contamination, we re-filtered haematoxylin and prepared fresh TBS-T, alcohol and H_2O_2 solutions every week. Fresh solutions were also

prepared before staining any *Vcan* and *Bcan* samples using the optimised protocol. We also included positive and negative controls to each staining batch, based on the requirements for each antibody.

Bibliography

Alberts, B. *et al.* (2017) *Molecular Biology of the Cell*, *6th edition, Garland Science*.

Alfaro-Almagro, F. *et al.* (2018) 'Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank', *NeuroImage*, 166, pp. 400-424.

Alyami, W., Kyme, A. and Bourne, R. (2022) 'Histological Validation of MRI: A Review of Challenges in Registration of Imaging and Whole-Mount Histopathology', *Journal of Magnetic Resonance Imaging*, 55(1), pp. 11–22.

Andersson, J.L.R., Skare, S. and Ashburner, J. (2003) 'How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging', *NeuroImage*, 20, pp. 870–888.

Andersson, J.L.R. and Sotiropoulos, S.N. (2016) 'An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging', *NeuroImage*, 125, pp. 1063–1078.

Bagnato, F. *et al.* (2018) 'Untangling the R2* contrast in multiple sclerosis: A combined MRI-histology study at 7.0 Tesla', *PLoS ONE*, 13(3), e0193839.

Basser, P.J., Mattiello, J. and LeBihan, D. (1994) 'MR diffusion tensor spectroscopy and imaging', *Biophysical Journal*, 66(1), pp. 259–267.

Bastiani, M. *et al.* (2017) 'Improved tractography using asymmetric fibre orientation distributions', *NeuroImage*, 158, pp. 205–218.

Batzdorf, C.S. *et al.* (2022) 'Sexual Dimorphism in Extracellular Matrix Composition and Viscoelasticity of the Healthy and Inflamed Mouse Brain', *Biology*, 11(2), p. 230.

Bautista, T. *et al.* (2021) 'Removal of Gibbs ringing artefacts for 3D acquisitions using subvoxel shifts', in *Proc. Intl. Soc. Mag. Reson. Med.*, p. 3535.

Bekku, Y. *et al.* (2009) 'Brevican distinctively assembles extracellular components at the large diameter nodes of Ranvier in the CNS', *Journal of Neurochemistry*, 108, pp. 1266–1276.

Bekku, Y. and Oohashi, T. (2011) 'Neurocan contributes to the molecular heterogeneity of the perinodal ECM', *Archives of Histology and Cytology*, 73(2), pp. 95–102.

Beller, J.A. and Snow, D.M. (2014) 'Proteoglycans: Road signs for neurite outgrowth', *Neural Regeneration Research*, 9(4), pp. 343–355.

Bogdan, R. *et al.* (2017) 'Imaging Genetics and Genomics in Psychiatry: A Critical Review of Progress and Potential', *Biological Psychiatry*, 82(3), pp. 165–175.

Brakebusch, C. *et al.* (2002) 'Brevican-Deficient Mice Display Impaired Hippocampal CA1 Long-Term Potentiation but Show No Obvious Deficits in Learning and Memory', *Molecular and Cellular Biology*, 22, pp. 7417–7427.

Breschi, A., Gingeras, T.R. and Guigo, R. (2019) 'Comparative transcriptomics in humans and mouse', *Nature Reviews Genetics*, 18, pp. 425–440.

Budde, M.D. and Annese, J. (2013) 'Quantifying anisotropy and fiber orientation in human brain histological sections', *Frontiers in Integrative Neuroscience*, 7.

Budde, M.D. and Frank, J.A. (2012) 'Examining brain microstructure using structure tensor analysis of histological sections', *NeuroImage*, 63, pp. 1–10.

Burton, P.R., Tobin, M.D. and Hopper, J.R. (2005) 'Key concepts in genetic epidemiology', *The Lancet*, 366, pp. 941–951.

Cahill, L.S. *et al.* (2012) 'Preparation of fixed mouse brains for MRI', *NeuroImage*, 60(2), pp. 933–939.

Chakravarty, M.M. *et al.* (2013) 'Performing label-fusion-based segmentation using multiple automatically generated templates', *Human Brain Mapping*, 34, pp. 2635–2654.

Chandran, P. *et al.* (2012) 'Magnetic resonance imaging and histological evidence for the blockade of cuprizone-induced demyelination in C57BL/6 mice', *Neuroscience*, 202, pp. 446–453.

Chang, E.H. *et al.* (2017) 'The role of myelination in measures of white matter integrity: Combination of diffusion tensor imaging and two-photon microscopy of CLARITY intact brains', *NeuroImage*, 147, pp. 253–261.

Cohen-Adad, J. (2014) 'What can we learn from T2* maps of the cortex?', *NeuroImage*, 93, pp. 189–200.

Crawley, J.N. (2008) 'Behavioral Phenotyping Strategies for Mutant Mice', *Neuron*, 57, pp. 809–818.

D'Arceuil, H.E., Westmoreland, S. and de Crespigny, A.J. (2007) 'An approach to high resolution diffusion tensor imaging in fixed primate brain', *NeuroImage*, 35, pp. 553–565.

Dauth, S. *et al.* (2016) 'Extracellular matrix protein expression is brain region dependent', *Journal of Comparative Neurology*, 524, pp. 1309–1336.

Denk, C. and Rauscher, A. (2010) 'Susceptibility weighted imaging with multiple echoes', *Journal of Magnetic Resonance Imaging*, 31(1), pp. 185–191.

Dours-Zimmermann, M.T. *et al.* (2009) 'Versican V2 assembles the extracellular matrix surrounding the nodes of Ranvier in the CNS', *Journal of Neuroscience*, 29(24), pp. 7731–7742.

Dyrby, T.B. *et al.* (2011) 'An ex-vivo imaging pipeline for producing high-quality and high-resolution diffusion-weighted imaging datasets', *Human Brain Mapping*, 32(4), pp. 544–563.

Einstein, A. (1905) 'Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen', *Annalen der Physik*, 322(8), pp. 549–560.

Elliott, L.T. *et al.* (2018) 'Genome-wide association studies of brain imaging phenotypes in UK Biobank', *Nature*, 562(7726), pp. 210–216.

Ermann, J. and Glimcher, L.H. (2012) 'After GWAS: Mice to the rescue?', *Current Opinion in Immunology*, 24, pp. 564–570.

Evanko, S.P. *et al.* (2018) 'The biochemistry and immunohistochemistry of versican', *Methods in Cell Biology*, 143, pp. 261–279.

Favuzzi, E. *et al.* (2017) 'Activity-Dependent Gating of Parvalbumin Interneuron Function by the Perineuronal Net Protein Brevican', *Neuron*, 95, pp. 639–655.

Fawcett, J.W., Oohashi, T. and Pizzorusso, T. (2019) 'The roles of perineuronal nets and the perinodal extracellular matrix in neuronal function', *Nature Reviews Neuroscience*, 20, pp. 451–465.

Ferrante, E. and Paragios, N. (2017) 'Slice-to-volume medical image registration: A survey', *Medical Image Analysis*, 39, pp. 101–123.

Foulcer, S.J., Day, A.J. and Apte, S.S. (2022) 'Isolation and Purification of Versican and Analysis of Versican Proteolysis', *Methods in Molecular Biology*, 2303, pp. 559–578.

Friedel, M. *et al.* (2014) 'Pydpiper: A flexible toolkit for constructing novel registration pipelines', *Frontiers in Neuroinformatics*, 8.

Ge, T., Schumann, G. and Feng, J. (2013) 'Imaging genetics — towards discovery neuroscience', *Quantitative Biology*, 1(4), pp. 227–245.

Geijs, D. *et al.* (2018) 'Automatic color unmixing of IHC stained whole slide images', SPIE Medical Imaging, Proceedings Volume 10581, p. 20.

Genovese, C.R., Lazar, N.A. and Nichols, T. (2002) 'Thresholding of statistical maps in functional neuroimaging using the false discovery rate', *NeuroImage*, 15(4), pp. 870–878.

Ghorbani, S. *et al.* (2022) 'Versican promotes T helper 17 cytotoxic inflammation and impedes oligodendrocyte precursor cell remyelination', *Nature Communications*, 13(1).

Ghorbani, S. and Yong, V.W. (2021) 'The extracellular matrix as modifier of neuroinflammation and remyelination in multiple sclerosis', *Brain*, 144(7), pp. 1958–1973.

Grussu, F. *et al.* (2017) 'Neurite dispersion: a new marker of multiple sclerosis spinal cord pathology?', *Annals of Clinical and Translational Neurology*, 4(9), pp. 663–679.

Gudbjartsson, H. and Patz, S. (1995) 'The rician distribution of noisy mri data', *Magnetic Resonance in Medicine*, 34(6), pp. 910–914.

Gurumurthy, C.B. and Kent Lloyd, K.C. (2019) 'Generating mouse models for biomedical research: Technological advances', *DMM Disease Models and Mechanisms*, 12 (1).

de Guzman, A.E. *et al.* (2016) 'Variations in post-perfusion immersion fixation and storage alter MRI measurements of mouse brain morphometry', *NeuroImage*, 142, pp. 687–695.

Haacke, E.M. *et al.* (2010) 'Susceptibility mapping as a means to visualize veins and quantify oxygen saturation', *Journal of Magnetic Resonance Imaging*, 32(3), pp. 663–676.

Haacke, E.M. *et al.* (2015) 'Quantitative susceptibility mapping: Current status and future directions', *Magnetic Resonance Imaging*, 33, pp. 1–25.

Hametner, S. *et al.* (2018) 'The influence of brain iron and myelin on magnetic susceptibility and effective transverse relaxation - A biochemical and histological validation study', *NeuroImage*, 179, pp. 117–133.

Hansen, C.B. *et al.* (2019) 'Characterization and correlation of signal drift in diffusion weighted MRI', *Magnetic Resonance Imaging*, 57, pp. 133–142.

Hanson, L.G. (2008) 'Is quantum mechanics necessary for understanding magnetic resonance?', *Concepts in Magnetic Resonance Part A: Bridging Education and Research*, 32(5), pp. 329–340.

Harlow, D.E. and Macklin, W.B. (2014) 'Inhibitors of myelination: ECM changes, CSPGs and PTPs', *Experimental Neurology*, 251, pp. 39–46.

Harten, I.A. *et al.* (2020) 'The synthesis and secretion of versican isoform V3 by mammalian cells: A role for N-linked glycosylation', *Matrix Biology*, 89, pp. 27–42.

He, T. *et al.* (2009) 'A phantom study of temperature-dependent MRI T2* measurement', *Journal of Cardiovascular Magnetic Resonance*, 11(S1).

Hernandez Cordero, A.I. *et al.* (2019) 'Genome-wide Associations Reveal Human-Mouse Genetic Convergence and Modifiers of Myogenesis, CPNE1 and STC2', *American Journal of Human Genetics*, 105, pp. 1222–1236.

Hernandez-Fernandez, M. *et al.* (2019) 'Using GPUs to accelerate computational diffusion MRI: From microstructure estimation to tractography and connectomes', *NeuroImage*, 105, pp. 1222–1236.

Hibar, D.P. *et al.* (2015) 'Common genetic variants influence human subcortical brain structures', *Nature*, 520, pp. 224–229.

Howard, A.F. *et al.* (2023) 'An open-source resource combining multi-contrast MRI and microscopy in the macaque brain', *Nature Communications*, 14.

Huang, S. *et al.* (2009) 'Manipulation of tissue contrast using contrast agents for enhanced MR microscopy in ex-vivo mouse brain', *NeuroImage*, 46, pp. 589–599.

Huang, S.Y. *et al.* (2015) 'The impact of gradient strength on in vivo diffusion MRI estimates of axon diameter', *NeuroImage*, 106, pp. 464–472.

Huszar, I.N. *et al.* (2023) 'Tensor image registration library: Deformable registration of stand-alone histology images to whole-brain post-mortem MRI data', *NeuroImage*, 265, p. 119792.

Iglesias, J.E. *et al.* (2018) 'Effect of fluorinert on the histological properties of formalin-fixed human brain tissue', *Journal of Neuropathology and Experimental Neurology*, 77, pp. 1085–1090.

Iozzo, R. V. (1998) 'Matrix proteoglycans: From molecular design to cellular function', *Annual Review of Biochemistry*, 67, pp. 609–652.

Jelescu, I. and Budde, M.D. (2017) 'Design and Validation of Diffusion MRI Models of White Matter', *Front Phys. Sec. Medical Physics and Imaging*, 5.

Jenkinson, M. *et al.* (2002) 'Improved optimization for the robust and accurate linear registration and motion correction of brain images', *NeuroImage*, 17(2), pp. 825–841.

Johansen-Berg, H. and Behrens, T.E.J. (2009) *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy*, *Diffusion MRI*. 2nd edition. Academic Press, Elsevier.

Johnson, G.A. *et al.* (2023) 'Merged magnetic resonance and light sheet microscopy of the whole mouse brain', *Proceedings of the National Academy of Sciences of the United States of America*, 120(17).

Jones, D.K. (2004) 'The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: A Monte Carlo study', *Magnetic Resonance in Medicine*, 51(4), pp. 807–815.

de Jong, J.M. *et al.* (2020) 'Remodeling of the interstitial extracellular matrix in white matter multiple sclerosis lesions: Implications for remyelination (failure)', *Journal of Neuroscience Research*, 98(7), pp. 1370–1397.

Kaden, E. *et al.* (2016) 'Multi-compartment microscopic diffusion imaging', *NeuroImage*, 139, pp. 346–359.

Kampourakis, K. (2017) *Making Sense of Genes*. Cambridge University Press.

Kellner, E. *et al.* (2016) 'Gibbs-ringing artifact removal based on local subvoxel-shifts', *Magnetic Resonance in Medicine*, 76, pp. 1574–1581.

Kor, D.Z.L. *et al.* (2022) 'An automated pipeline for extracting histological stain area fraction for voxelwise quantitative MRI-histology comparisons', *NeuroImage*, 264, p. 119726.

Lau, L.W. *et al.* (2012) 'Chondroitin sulfate proteoglycans in demyelinated lesions impair remyelination', *Annals of Neurology*, 72(3), pp. 419–432.

Lau, L.W. *et al.* (2013) 'Pathophysiology of the brain extracellular matrix: A new target for remyelination', *Nature Reviews Neuroscience*, 14(10), pp. 722–729.

Lauterbur, P.C. (1973) 'Image formation by induced local interactions: Examples employing nuclear magnetic resonance', *Nature*, 242(5394), pp. 190–191.

Lazari, A. and Lipp, I. (2021) 'Can MRI measure myelin? Systematic review, qualitative assessment, and meta-analysis of studies validating microstructural imaging with myelin histology', *NeuroImage*, 230, p. 117744.

Lee, J. *et al.* (2011) 'T2*-based fiber orientation mapping', *NeuroImage*, 57(1), pp. 225–234.

Lerch, J.P., Sled, J.G. and Henkelman, R.M. (2010) 'MRI phenotyping of genetically altered mice', *Magnetic Resonance Neuroimaging, Methods in Molecular Biology, vol 711,* Humana Press, pp. 349-361.

Li, W. *et al.* (2015) 'A method for estimating and removing streaking artifacts in quantitative susceptibility mapping', *NeuroImage*, 108, pp. 111–122.

Li, W., Wu, B. and Liu, C. (2011) 'Quantitative susceptibility mapping of human brain reflects spatial variation in tissue composition', *NeuroImage*, 55, pp. 1645–1656.

Liu, C. *et al.* (2011) 'High-field (9.4T) MRI of brain dysmyelination by quantitative mapping of magnetic susceptibility', *NeuroImage*, 56(3), pp. 930–938.

Ma, R. *et al.* (2020) 'A field-monitoring-based approach for correcting eddy-current-induced artifacts of up to the 2nd spatial order in human-connectome-project-style multiband diffusion MRI experiment at 7T: A pilot study', *NeuroImage*, 216, p. 116861.

Mague, S.D. *et al.* (2009) 'Mouse model of OPRM1 (A118G) polymorphism has sex-specific effects on drug-mediated behavior', *Proceedings of the National Academy of Sciences of the United States of America*, 106, pp. 10847–10852.

Mahajan, R. (2018) 'Chondroitinase ABC Enzyme: A Potential Treatment Option for Spinal Cord Injury', *International Journal of Applied and Basic Medical Research*, 8, p. 203.

Mann, C. *et al.* (2018) 'The effect of age on vertex-based measures of the grey-white matter tissue contrast in autism spectrum disorder', *Molecular Autism*, 9(1).

Manolio, T.A. *et al.* (2009) 'Finding the missing heritability of complex diseases', *Nature*, 461(7265), pp. 747–753.

Mauvais-Jarvis, F., Arnold, A.P. and Reue, K. (2017) 'A Guide for the Design of Pre-clinical Studies on Sex Differences in Metabolism', *Cell Metabolism*, 25(6), pp. 1216–1230.

Milev, P. *et al.* (1998) 'Differential regulation of expression of hyaluronan-binding proteoglycans in developing brain: Aggrecan, versican, neurocan, and brevican', *Biochemical and Biophysical Research Communications*, 247(2), pp. 207–212.

Mjaatvedt, C.H. *et al.* (1998) 'The Cspg2 gene, disrupted in the hdf mutant, is required for right cardiac chamber and endocardial cushion formation', *Developmental Biology*, 202, pp. 56–66.

Mollink, J. *et al.* (2019) 'White matter changes in the perforant path area in patients with amyotrophic lateral sclerosis', *Neuropathology and Applied Neurobiology*, 45(6), pp. 570–585.

Nandadasa, S. *et al.* (2021) 'A new mouse mutant with cleavage-resistant versican and isoform-specific versican mutants demonstrate that proteolysis at the Glu441-Ala442 peptide bond in the V1 isoform is essential for interdigital web regression', *Matrix Biology Plus*, 10, p. 100064.

Nica, A.C. and Dermitzakis, E.T. (2013) 'Expression quantitative trait loci: present and future', *Philosophical Transactions of the Royal Society B*, 368, p. 20120362.

Nicholson, C. and Syková, E. (1998) 'Extracellular space structure revealed by diffusion analysis', *Trends in Neurosciences*, 21(5), pp. 207–215.

Niederöst, B.P. *et al.* (1999) 'Bovine CNS myelin contains neurite growth-inhibitory activity associated with chondroitin sulfate proteoglycans', *Journal of Neuroscience*, 19(20), pp. 8979–8989.

Nishimura, D. (1996) *Principles of Magnetic Resonance Imaging*. Stanford University.

Ogawa, T. *et al.* (2001) 'Brevican in the developing hippocampal fimbria: Differential expression in myelinating oligodendrocytes and adult astrocytes suggests a dual role for brevican in central nervous system fiber tract development', *Journal of Comparative Neurology*, 432(3), pp. 285–295.

Palombo, M. *et al.* (2020) 'SANDI: A compartment-based model for non-invasive apparent soma and neurite imaging by diffusion MRI', *NeuroImage*, 215, p. 116835.

Perissinotto, D. *et al.* (2000) 'Avian neural crest cell migration is diversely regulated by the two major hyaluronan-binding proteoglycans PG-M/versican and aggrecan', *Development*, 127(13), pp. 2823–2842.

Robinson, S.D. *et al.* (2017) 'Combining phase images from array coils using a short echo time reference scan (COMPOSER)', *Magnetic Resonance in Medicine*, 77, pp. 318–327.

Rodgers, G. *et al.* (2022) 'Virtual histology of an entire mouse brain from formalin fixation to paraffin embedding. Part 2: Volumetric strain fields and local contrast changes', *Journal of Neuroscience Methods*, 365, p. 109385.

Roebroeck, A., Miller, K.L. and Aggarwal, M. (2019) 'Ex-vivo diffusion MRI of the human brain: Technical challenges and recent advances', *NMR in Biomedicine*, 32(4).

Ruifrok, A.C. and Johnston, D.A. (2001) 'Quantification of histochemical staining by color deconvolution', *Analytical and Quantitative Cytology and Histology*, 23(4), pp. 291–299.

Salat, D.H. *et al.* (2011) 'Hippocampal degeneration is associated with temporal and limbic gray matter/white matter tissue contrast in Alzheimer's disease', *NeuroImage*, 54(3), pp. 1795–1802.

Scelsi, M.A. (2020) *The genetics of dementia progression and heterogeneity*. University College London.

Schilling, K.G. *et al.* (2022) 'Recommendations and guidelines from the ISMRM Diffusion Study Group for preclinical diffusion MRI: Part 2 — Ex-vivo imaging', *arXiv2209.13371*.

Schmalfeldt, M. *et al.* (2000) 'Brain derived versican V2 is a potent inhibitor of axonal growth', *Journal of Cell Science*, 113(5), pp. 807–816.

Schurr, R. and Mezer, A.A. (2021) 'The glial framework reveals white matter fiber architecture in human and primate brains', *Science*, 374(6568).

Schweser, F. *et al.* (2011) 'Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: An approach to in vivo brain iron metabolism?', *NeuroImage*. 54, pp. 2789–2807.

Seehaus, A. *et al.* (2015) 'Histological validation of high-resolution DTI in human post-mortem tissue', *Frontiers in Neuroanatomy*, 9.

Seiberlich, N. *et al.* (eds) (2020) *Quantitative Magnetic Resonance Imaging*. CA Academic Press.

Seidenbecher, C.I. *et al.* (1995) 'Brevican, a chondroitin sulfate proteoglycan of rat brain, occurs as secreted and cell surface glycosylphosphatidylinositol-anchored isoforms', *Journal of Biological Chemistry*, 270(45), pp. 27206–27212.

Shakhnovich, V. (2018) 'It's Time to Reverse our Thinking: The Reverse Translation Research Paradigm', *Clinical and Translational Science*, 11(2), pp. 98–99.

Slatkin, M. (2008) 'Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future', *Nature Reviews Genetics*, 9(6), pp. 477–485.

Smart, A. *et al.* (2023) 'Protocol for tissue processing and paraffin embedding of mouse brains following ex vivo MRI', *STAR Protocols*, 4(4), p. 102681.

Smith, S.M. *et al.* (2004) 'Advances in functional and structural MR image analysis and implementation as FSL', in *NeuroImage*. 23, pp. S208–S219.

Smith, S.M. *et al.* (2006) 'Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data', *NeuroImage*. 4, pp. 1487-1505.

Smith, S.M. *et al.* (2021) 'An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank', *Nature Neuroscience*. 24, pp. 737-745.

Snow, D.M., Mullins, N. and Hynds, D.L. (2001) 'Nervous system-derived chondroitin sulfate proteoglycans regulate growth cone morphology and inhibit neurite outgrowth: A light, epifluorescence, and electron microscopy study', *Microscopy Research and Technique*, 54(5), pp. 273–286.

Song, I. and Dityatev, A. (2018) 'Crosstalk between glia, extracellular matrix and neurons', *Brain Research Bulletin*, 136, pp. 101–108.

Soria, F.N. *et al.* (2020) 'Current Techniques for Investigating the Brain Extracellular Space', *Frontiers in Neuroscience*, 14.

Stadhouders, R. *et al.* (2012) 'Transcription regulation by distal enhancers', *Transcription*, 3(4), pp. 181–186.

Stephens, E.H., Chu, C.K. and Grande-Allen, K.J. (2008) 'Valve proteoglycan content and glycosaminoglycan fine structure are unique to microstructure, mechanical load and age: Relevance to an age-specific tissue-engineered heart valve', *Acta Biomaterialia*, 4(5), pp. 1148–1160.

Stephenson, E.L. and Yong, V.W. (2018) 'Pro-inflammatory roles of chondroitin sulfate proteoglycans in disorders of the central nervous system', *Matrix Biology*, 71–72, pp. 432–442.

Sudlow, C. *et al.* (2015) 'UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age', *PLoS Medicine*, 12(3).

Sun, S.W. *et al.* (2005) 'Formalin fixation alters water diffusion coefficient magnitude but not anisotropy in infarcted brain', *Magnetic Resonance in Medicine*, 53(6), pp. 1447–1451.

Sundgren, P.C. *et al.* (2004) 'Diffusion tensor imaging of the brain: Review of clinical applications', *Neuroradiology*, 46(5), pp. 339–350.

Sur, I.K. *et al.* (2012) 'Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors', *Science*, 338, pp. 1360–1363.

Suvarna, S.K., Layton, C. and Bancroft, J.D. (2013) *Bancroft's Theory and Practice of Histological Tehchniques (Seventh Edition)*. Elsevier, Churchill Livingstone.

Svenson, K.L. *et al.* (2007) 'Multiple trait measurements in 43 inbred mouse strains capture the phenotypic diversity characteristic of human populations', *Journal of Applied Physiology*, 102, pp. 2369–2378.

Tam, V. *et al.* (2019) 'Benefits and limitations of genome-wide association studies', *Nature Reviews Genetics*, 20, pp. 467–484.

Tanner, J.E. and Stejskal, E.O. (1968) 'Restricted self-diffusion of protons in colloidal systems by the pulsed-gradient, spin-echo method', *The Journal of Chemical Physics*, 49(4), pp. 1768–1777.

Tax, C.M.W. *et al.* (2022) 'What's new and what's next in diffusion MRI preprocessing', *NeuroImage*, 249, p. 118830.

Tendler, B.C. *et al.* (2021) 'A method to remove the influence of fixative concentration on postmortem T2 maps using a kinetic tensor model', *Human Brain Mapping*, 42(18), pp. 5956–5972.

Tendler, B.C. *et al.* (2022) 'The Digital Brain Bank, an open access platform for post-mortem imaging datasets', *eLife*, 11, p. e73153.

Tisca, C. *et al.* (2020) 'Comparison of NODDI parameter estimates in staggered versus non-staggered two-shell acquisitions', in *ESMRMB*.

Ueno, H. *et al.* (2017) 'Parvalbumin neurons and perineuronal nets in the mouse prefrontal cortex', *Neuroscience*, 343, pp. 115–127.

Uffelmann, E. *et al.* (2021) 'Genome-wide association studies', *Nature Reviews Methods Primers*, 1(1).

Wang, C. *et al.* (2020) 'Methods for quantitative susceptibility and R2* mapping in whole post-mortem brains at 7T applied to amyotrophic lateral sclerosis', *NeuroImage*, 222(May), p. 117216.

Wang, C. *et al.* (2022) 'Phenotypic and genetic associations of quantitative magnetic susceptibility in UK Biobank brain imaging', *Nature Neuroscience*, 25(6), pp. 818–831.

Wang, Q. *et al.* (2020) 'The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas', *Cell*, 181(4), pp. 936-953.

Wansapura, J.P. *et al.* (1999) 'NMR relaxation times in the human brain at 3.0 Tesla', *Journal of Magnetic Resonance Imaging*, 9(4), pp. 531–538.

Waterston, R.H. *et al.* (2002) 'Initial sequencing and comparative analysis of the mouse genome', *Nature*, 420, pp. 520–562.

Wei, H. *et al.* (2015) 'Streaking artifact reduction for quantitative susceptibility mapping of sources with large dynamic range', *NMR in Biomedicine*, 28, pp. 1294-1303.

van der Weijden, C.W.J. *et al.* (2021) 'Myelin quantification with MRI: A systematic review of accuracy and reproducibility', *NeuroImage*, 226, p. 117561.

Wight, T.N. *et al.* (2014) 'Versican and the regulation of cell phenotype in disease', *Biochimica et Biophysica Acta - General Subjects*, 1840(8), pp. 2441–2451.

Wight, T.N. *et al.* (2020) 'Versican—A Critical Extracellular Matrix Regulator of Immunity and Inflammation', *Frontiers in Immunology*, 11.

Winkler, A.M. *et al.* (2014) 'Permutation inference for the general linear model', *NeuroImage*, 92, pp. 381–397.

Xin, Y. *et al.* (2022) 'A review of imaging genetics in Alzheimer's disease', *Journal of Clinical Neuroscience*, 100, pp. 155–163.

Xu, T. *et al.* (2018) 'The effect of realistic geometries on the susceptibility-weighted MR signal in white matter', *Magnetic Resonance in Medicine*, 79, pp. 489-500.

associates with astrocytes ensheathing cerebellar glomeruli and inhibits neurite outgrowth from granule neurons', *Journal of Neuroscience*, 17(20), pp. 7784–7795.

Yee, Y. *et al.* (2018) 'Structural covariance of brain region volumes is associated with both structural connectivity and transcriptomic similarity', *NeuroImage*, 179, pp. 357–372.

Yong-Hing, C.J. *et al.* (2005) 'Magnetic resonance imaging and mathematical modeling of progressive formalin fixation of the human brain', *Magnetic Resonance in Medicine*, 54(2), pp. 324–332.

Zhang, H. *et al.* (2012) 'NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain', *NeuroImage*, 61, pp. 1000-1016.

Zheng, P.S. *et al.* (2004) 'Versican/PG-M G3 domain promotes tumor growth and angiogenesis.', *The FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 18(6), pp. 754–756.

Zhou, Z. *et al.* (2016) 'Cerebral cavernous malformations arise from endothelial gain of MEKK3-KLF2/4 signalling', *Nature*, 532(7597), pp. 122–126.